The Dissertation Committee for James Hansen Delfeld
certifies that this is the approved version of the following dissertation:

Labeling and denoising geometrically parameterized data with applications to Cryo-EM

Committee:

Ronny Hadani, Supervisor

Chandrajit Bajaj

Amit Singer

Rachel Ward

Peter Müeller

Labeling and denoising geometrically parameterized data with applications to Cryo-EM

by

James Hansen Delfeld, B.S. ; B.A.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2015

Labeling and denoising geometrically parameterized data with applications to Cryo-EM

Publication No. _____

James Hansen Delfeld, Ph.D.
The University of Texas at Austin, 2015

Supervisor: Ronny Hadani

Many data sets encountered in practice depend continuously on a geometric parameter. An important example of this is image collections from Cryo-EM experiments, where the images depend continuously on the orientation of molecules. The first part of the thesis considers the problem of labeling geometrically parameterized data sets. It shows that for this problem the popular method of spectral clustering is sensitive to noise. It then presents a categorical optimization solution which is unbiased and robust to noise. The second part of the thesis presents a method to denoise collections of Cryo-EM images. The method represents Cryo-EM image collections as a single vector in a high dimensional space, and computes a low dimensional subspace which well contains the signal of the vector. By projecting the vector of images onto this subspace, the image collection is denoised. The thesis shows that the output images are centered, and that their SNR grows linearly with the number of input images.

Table of Contents

List of Tables

List of Figures

Part I

Labeling geometrically parameterized data

Chapter 1

Introduction

The first part of this thesis addresses the problem of labeling a collection of data points. The data is assumed to have two structures:

- **An unknown intrinsic classification**. By intrinsic classification we mean the data can be naturally segmented into a discrete number of classes.
- **A known geometric parameterization**. By geometric parameterization we mean the values of the data points depend continuously on a metric space parameter.

Before giving formal definitions for classification and geometric parameterization we illustrate them with examples.

- **Feature Tracking** Over an interval of time a sequence of images are taken of a dynamic scene. For each image, features are extracted. The goal of feature tracking is to match the features across all the images [25]. In this problem the data is the collection of extracted features, the geometric parameterization is time, the intrinsic classification is feature segmentation by objects (cars, people, buildings, etc...) in the scene.

- **Heterogeneity in Cryo-EM** Cryo Electron Microscopy (Cryo-EM) is a technique to find the potential function of biomolecules. It works by freezing in ice a large collection of randomly oriented copies of a single molecule (on the order of 10,000) and imaging them with an electron beam. The resulting 2D projection images are then used to form a 3D reconstruction of the molecule's potential function [6]. The problem of heterogeneity happens when there is a mixture of two or more molecules in the ice. If standard reconstruction techniques are used the resulting reconstruction will be an uninformative average of the different molecule types found in the sample [14]. Suppose an algorithm determines the molecular orientation associated with every image, then a solution to the heterogeneous problem would partition the images by molecule type so separate reconstructions can be performed. In this problem the data is the collection of projection images, the geometric parameterization is the space of orientations, the intrinsic classification is image classification by molecule type.

- **Image Segmentation** Image segmentation is the problem of extracting different objects found in an image [22]. In this problem the data is the intensities of each pixel, the geometric parameterization is the set of pixel locations, the intrinsic classification is pixel segmentation by objects found in the image.

We proceed to give formal definitions of classification and parameterization.

Definition 1. *Given a data set D, a **classification** consists of a pair (C, c) where C is a finite set and c is a map $c : D \to C$.*

Definition 2. *Given a data set D, a* ***geometric parameterization*** *consists of a pair (X, x) where X is a finite space equipped with a metric m_X and x is a map $x : D \to X$.*

Applying these definitions to the above examples gives the following table.

Example	D	X	C
Feature Tracking	Features	Time	Objects in scene
Cryo-EM	Images	Orientations	Molecule types
Image Segmentation	Pixel Int.	Pixel Loc.	Objects in image

Table 1.1: Examples of classification and geometric parameterization.

The main assumption of the thesis is that given a data set D equipped with a classification (C, c), there is a method for taking certain pairs of data points and deciding if they belong to the same class. To formalize this assumption we introduce the notion of *affinity kernel*.

Definition 3. *The* ***affinity kernel****, δ_c, is a map $\delta_c : D \times D \to \mathbb{R}$ defined by:*

$$\delta_c(d, d') = \begin{cases} 1 & \text{if} \quad c(d) = c(d') \\ 0 & \text{if} \quad c(d) \neq c(d'). \end{cases}$$

The difficulty is that in practice one does not have access to the clean affinity kernel, $\delta_c(d, d')$, but rather to a noisy version of it:

$$\widehat{\delta}_c(d, d') = \delta_c(d, d') + \epsilon(d, d'),$$

where $\epsilon(d, d')$ is noise (models of the noise will be described later). We refer to $\widehat{\delta}_c(d, d')$ as the empirical affinity kernel. We can now give the formal definition of the labeling problem.

Labeling Problem *Suppose D is a data set equipped with an unknown classification (C, c). Given as input a complete collection of empirical affinity kernels, $\widehat{\delta}_c(d, d')$ for all pairs $(d, d') \in D \times D$. Produce as output a labeling of D consistent with the classification C, i.e. two data points have the same label if and only if they belong to the same class.*

Remark: The labeling problem makes two strong assumptions. The first assumption is that the data has an intrinsic classification. The second assumption is that the affinity kernels give a hard, 0 or 1, characterization about the class relationship between data points. This should be contrasted with standard clustering formulations which typically takes as input similarity measurements between data points and outputs a partition of the data so that similarities are large between data points belonging to the same partition set.

We proceed to discuss the problem of labeling geometrically parameterized data and explain how it contrasts with the labeling problem defined above. The main difficulty when labeling geometrically parameterized data, is that data point variation due to difference in geometric parameter will often overwhelm data point variation due to difference in class. The following numerical experiment motivated by the heterogeneity problem in Cryo-EM illustrates this phenomena. Figure 1.1 shows two phantoms where the left phantom differs from the right phantom only by the presence of a single ball. This serves as an abstraction of the subtle differences between molecule types encountered in heterogeneous Cryo-EM. Figure 1.2 shows projection images of the two phantoms taken from the same orientation. One can see that the im-

ages are visually nearly indistinguishable. Figure 7.1 shows projection images of phantoms taken from different orientations. One can see that the images are radically different *even if they came from the same phantom type*. To conclude, we see from this example that image variation due to difference in orientation overwhelms image variation due to difference in phantom type.

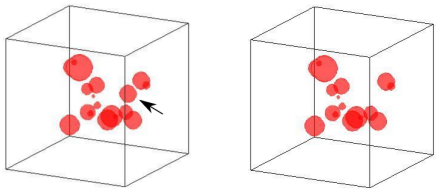


Figure 1.1: The two generated phantoms. The presence of a single extra ball distinguishes the left phantom from the right phantom.

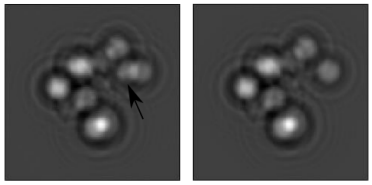
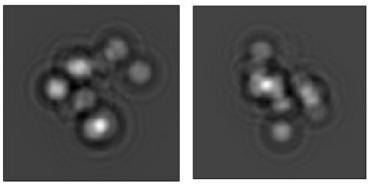
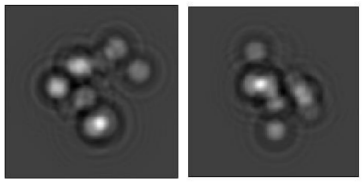


Figure 1.2: Projection images of the two phantoms taken from the same orientation. The arrow points at the difference in the images.



(a) Different phantoms.



(b) Same phantom.

Figure 1.3: Projection images from different orientations of the phantoms. The left pair of images came from different phantoms. The right pair of images came from the same phantom.

Let us return to the general problem of labeling geometrically parameterized data where we assume that data point variation due to geometric parameter is much larger then data point variation due to difference in class. First, suppose we observe a pair of data points with largely different geometric parameters. The two data points will look very different regardless of whether or not they belong to the same class. Therefore a faithful empirical affinity kernel cannot be computed for this pair. Next suppose we observe a pair of data points with similar geometric parameters. For this pair, variation due to difference in geometric parameter is minimal and any observed variation can be attributed mainly to difference in class. Therefore a faithful empirical affinity kernel can be computed for this pair. We can now describe the formal problem of labeling geometrically parameterized data.

Problem of labeling parameterized data *Suppose D is a data set equipped with an unknown classification (C, c) and a known geometric parameterization (X, x). Fix a constant radius $r > 0$. Given as input empirical affinity kernels between all pairs of data points whose geometric parameters are within distance r of each other,*

$$\left\{\widehat{\delta}_c(d, d') : m_X(x(d), x(d')) \leq r\right\}.$$

Produce as output a labeling of D consistent with the C classification, i.e. two data points have the same label if and only if they belong to the same class.

1.1 Previous Work

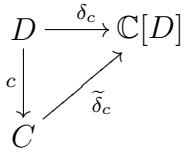
Spectral clustering is currently the most commonly used method to solve the labeling problem. For the sake of completeness, we give a brief outline of this method and explain it's inadequacy to solve the labeling problem

for geometrically parameterized data.

First let us consider the case when affinity kernels are known between every pair of data points. The affinity kernel can be used to embed the data set D into the vector space $\mathbb{C}[D]$ through the map:

$$\delta_c : D \to \mathbb{C}[D],$$

given by $[\delta_c(d)](d') = \delta_c(d, d')$. We note that δ_c is constant on the fibers of c, that is, data points belonging to the same class get mapped to the same vector. Therefore, δ_c defines a map $\widetilde{\delta}_c$ which fits in the following commutative diagram:



this means that, $\widetilde{\delta}_c(C) = \delta_c(D)$. Applying principal component analysis to the collection of vectors $\delta_c(D)$ one recovers the subspace spanned by the vectors $\widetilde{\delta}_c(C)$. Projecting the vectors in $\delta_c(D)$ onto this subspace reduces the dimension from $|D|$ to $|C|$. Finally, applying k-means clustering on these projected vectors establishes a labeling of the data. It can be shown that these labels are consistent with the classification (C, c) [18]. However, in practice we do not have access to the clean affinity kernels, but rather to their noisy approximations in the form of empirical affinity kernels. It can be shown that when given a complete set of empirical affinity kernels, the method of spectral clustering is robust to noise [29].

We now consider the method of spectral clustering when applied to

data equipped with a geometric parameterization (X, x). Recall, that for geometrically paramerized data we only have access to affinity kernels locally, that is between pairs of data points close to each other. The conventional spectral clustering method proceeds as before, taking the top eigenvectors of the local affinity kernel matrix and applying k-means clustering. In the absence of noise one can show that the method exactly recovers a correct labeling of the data points. However, in the presence of noise the local empirical affinity kernel matrix losses information about class membership and begins to resemble the Laplacian on X. As a result, applying k-means clustering to the top eigenvectors segments the data by geometry as opposed to class membership. This phenomena is quantified in section 3 and is the main motivation for the development of this thesis. *In summary: having access to empirical affinity kernels only locally causes the spectral clustering method to become sensitive to noise.*

1.2 Main Results

In the first part of this thesis we develop the framework of categorical optimization in the context of the labeling problem. In a nutshell, the paradigm shift is that standard optimization methods solves problems by defining a parameter space Ω and characterizes the solution as $\text{argmax}_{\omega\in\Omega} F(\omega)$ for some scoring function $F : \Omega \to \mathbb{R}$. In contrast, categorical optimization solves problems by defining a category $\mathcal{C}$ and characterizes solutions as an object in this category [10] [11].

- The first result of the thesis is a categorical optimization solution to the labeling problem for unparameterized data. We show that, in this

simple case, the performance of the solution is equivalent to the method of spectral clustering.

- The second result of the thesis is a theorem quantifying the noise sensitivity of the spectral clustering method when it is applied to geometrically parameterized data.

- The third result of the thesis is a categorical optimization solution to the labeling problem for geometrically parameterized data, which is unbiased and robust to noise.

1.3 Structure of part I

The first part of the thesis consists of five chapters and two appendices.

- **Chapter 2**: In this chapter we introduce the category of admissible labelings, and show that solving the labeling problem is equivalent to computing an object in this category. We then introduce the dual category of admissible algebra labelings. Finally we show how the graph Laplacian can be used to compute a particular admissible algebra labeling in a robust manner.

- **Chapter 3**: In this chapter we consider the spectral clustering method applied to geometrically parameterized data. We first explain that in the absence of noise applying k-means to the spectrum of a local graph Laplacian gives a precise solution. We then show that in the presence of noise the spectrum becomes biased to geometry. The bias to geometry is quantified and illustrated via a numerical experiment.

- **Chapter 4**: In this chapter we present a categorical optimization solution to the labeling problem for geometrically parameterized data. The solution constructs labelings on small patches of the data set, then combines these local labelings into a global labeling. This is formalized by the category of admissible partition labelings. Next the category is dualized giving the category of admissible partition algebras. Finally an operator is introduced which can be computed from the data and whose spectrum recovers, in a robust manner, a particular admissible partition algebra.

- **Chapter 5**: In this chapter we present numerical results from applying the categorical optimization solution of chapter 3 to the problem of Cryo-EM heterogeneity.

- **Chapter 6**: In this chapter we give a summary and conclusion for the first part of the thesis. We also outline how the tools developed in the first part of the thesis could be applied to more general classification problems.

- **Appendix A**: Gives an overview of the duality between finite sets and finite complex algebras.

- **Appendix B**: Contains proofs of theorems and propositions presented in the first part of thesis.

Chapter 2

Categorical solution to the labeling problem

2.1 Categorical formulation of the labeling problem

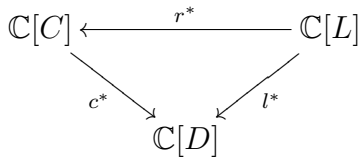
A solution to the labeling problem is a labeling of D consistent with the classification (C, c) i.e. two data points have the same label if and only if they belong to the same class. We call such labelings *admissible*. Formally a labeling of D is a pair (L, l) with L a finite set and l a map $l : D \to L$. The admissible criteria can be expressed using a commutative diagram.

Definition 4. *A labeling (L, l) is called an **admissible labeling** if there exists a bijection (relabeling) $r : C \to L$ fitting in the following commutative diagram:*

$$\begin{array}{ccc} C & \xrightarrow{\quad r \quad} & L \\ & {\scriptstyle c}\nwarrow \quad \nearrow{\scriptstyle l} & \\ & D & \end{array} \tag{2.1}$$

Note that there is no preferred admissible labeling. Rather, the collection of all admissible labelings form a category, where all objects are isomorphic to one another in a unique manner. A categorical solution to the labeling problem amounts to constructing an object in this category. Invoking the standard duality between finite sets and finite algebras (reviewed in Appendix A) we replace the sets C, D, and L with the corresponding algebras of complex valued functions $\mathbb{C}[C]$, $\mathbb{C}[D]$, and $\mathbb{C}[L]$ and the maps l, c, r with the pullback algebra morphisms l^*, c^*, r^*, respectively. In dual terms diagram

(8.1) translates to the following commutative diagram:



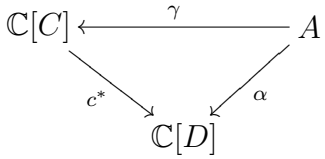
(2.2)

The commutativity follows from the functorality of pullbacks. Specifically, since $l = r \circ c$, taking pullbacks gives:

$$l^* = (r \circ c)^* = c^* \circ r^*.$$

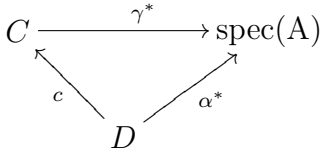
From a computational prospective diagram (2.2) is more appealing then diagram (8.1) because vector spaces and linear maps can be calculated using the powerful tools of linear algebra. This motivates the following definition.

Definition 5. *An **admissible algebra labeling** is a pair (A, α) where A is a finite algebra and α is an algebra morphism $\alpha : A \to \mathbb{C}[D]$ such that there exits an algebra isomorphism $\gamma : A \to \mathbb{C}[C]$ fitting in the following commutative diagram:*



(2.3)

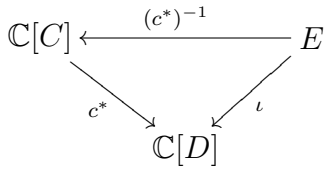
Given an admissible algebra labeling (A, α), the dual pair $(\text{spec}(A), \alpha^*)$ is an admissible labeling. To see this take the dual of diagram (2.3):



because diagram (2.3) commutes, the functorality of pullbacks implies that this diagram commutes as well. Therefore, we conclude that $(\text{spec}(A), \alpha^*)$ is an admissable labeling.

2.2 Construction of an admissible algebra labeling

In this section we explain how to use the affinity kernels to construct a particular admissible algebra labeling. Recall that the map $c^* : \mathbb{C}[C] \to \mathbb{C}[D]$ is an injective algebra morphism. Denote the image of c^* by E. The algebra E, equipped with the canonical embedding map $\iota : E \hookrightarrow \mathbb{C}[D]$ forms an admissible algebra labeling, (E, ι). To see this, verify that the following diagram commutes:



Once E is computed, the map ι is given for free. Therefore, the labeling problem can be reduced to computing the subalgebra $E \subset \mathbb{C}[D]$. The upshot is that the algebra E can be recovered from the spectrum of a specific operator, $\mathcal{G} : \mathbb{C}[D] \to \mathbb{C}[D]$, defined in terms of the affinity kernels. The operator $\mathcal{G}$ is called the (unnormalized) graph Laplacian and is defined by:

$$(\mathcal{G}f)(d) = \sum_{d' \in D} \delta_c(d, d') f(d'),$$

for all $f \in \mathbb{C}[D]$. The precise relationship between the spectrum of $\mathcal{G}$ and the algebra E is the content of the following theorem.

Theorem 1. ***(Spectrum of the graph Laplacian)***

1. $\mathrm{Im}(\mathcal{G}) = E$.

2. *The non-zero eigenvalues of* $\mathcal{G}$ *are the number of data points in each class,* $|c^{-1}(a)|$ *for all* $a \in C$*, with corresponding eigenspaces* $\mathrm{span}(c^*(\chi_a))$ *where* χ_a *is the characteristic function of* a*.*

For the proof see Appendix B. To summarize: the affinity kernels can be used to construct the graph Laplacian operator, $\mathcal{G}$. The spectrum of $\mathcal{G}$ recovers E which equipped with the canonical embedding map $\iota : E \hookrightarrow \mathbb{C}[D]$ gives an admissible algebra labeling, thus solving the labeling problem.

2.3 Unbiasedness of the graph Laplacian

As explained earlier, in practice one does not have access to the affinity kernels, $\delta_c(d, d')$, but rather to a collection of empirical affinity kernels:

$$\widehat{\delta}_c(d, d') = \delta_c(d, d') + \epsilon(d, d'),$$

which define an empirical graph Laplacian operator, $\widehat{\mathcal{G}}$. In this section we show that the spectrum of $\widehat{\mathcal{G}}$ is robust to the presence of the noise term, $\epsilon(d, d')$. As a model of noise we first use random false positives and false negatives, that is, with probability p a pair of data points are wrongly identified to be in the same class false positives) and with probability q a pair of data points in the same class are not identified as such (false negatives). This error model yields the following stochastic kernel:

$$\begin{aligned}\mathbb{P}(\widehat{\delta}_c(d, d') = 1 | \delta_c(d, d') = 0) &= p \\ \mathbb{P}(\widehat{\delta}_c(d, d') = 0 | \delta_c(d, d') = 1) &= q,\end{aligned}$$

for all $(d, d') \in D \times D$. Taking the expectation of $\widehat{\mathcal{G}}$ under this error model we get:

$$\begin{aligned}(\mathbb{E}\widehat{\mathcal{G}}f)(d) &= \sum_{d' \in D} [(1-q)\delta_c(d, d') + p(1 - \delta_c(d, d'))] f(d') \\ &= \sum_{d' \in D} [(1-q-p)\delta_c(d, d') + p] f(d') \\ &= (1-q-p)(\mathcal{G}f)(d) + p \sum_{d' \in D} f(d'),\end{aligned} \tag{2.4}$$

for all $f \in \mathbb{C}[D]$. Equation (2.4) expresses $\mathbb{E}\widehat{\mathcal{G}}$ as a sum of two operators. The first term is proportional to the clean graph Laplacian and the second term is proportional to the averaging operator $\mathcal{A} : \mathbb{C}[D] \to \mathbb{C}[D]$ defined by:

$$(\mathcal{A}f)(d) = \sum_{d' \in D} f(d'),$$

for all $f \in \mathbb{C}[D]$. The averaging operator surjects into the one-dimensional vector space of constant functions which fortunately is contained in E. Therefore, the spectrum of $\mathbb{E}\widehat{\mathcal{G}}$ recovers E.

Theorem 2. *(**Spectrum of the expected graph Laplacian**) We have:*

$$\mathrm{Im}(\mathbb{E}\widehat{\mathcal{G}}) = c^*\mathbb{C}[C]$$

For the proof see Appendix B. Note that we can generalize the error model to any model which gives an expected empirical kernel of the form:

$$\mathbb{E}\delta_c(d, d') = p\delta_c(d, d') + q, \tag{2.5}$$

where p and q are fixed constants. This includes, for example, empirical kernels taking soft (not 0 or 1) values, as long as the expected kernel is given by equation (2.5).

2.4 Numerical Dualization

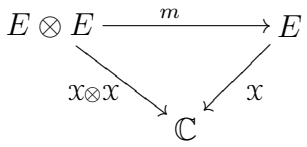
Given an empirical graph Laplacian $\widehat{\mathcal{G}}$, the span of it's top $|C|$ eigenvectors define a subspace denoted by $\widehat{E}$. As a vector space, $\widehat{E}$ is close to the algebra E. This follows from theorem 2 and standard results from random matrix theory. In this section we explain how $\widehat{E}$ can be used to construct a labeling of D.

The main difficulty is that although $\widehat{E}$ is close to E there is no guarantee that $\widehat{E}$ is an algebra. Therefore, we cannot invoke algebraic duality to produce a labeling. Instead, we introduce a procedure called "numerical dualization" that yields:

1. A finite set $\mathrm{nspec}(\widehat{E})$, called the *numerical spectrum* of $\widehat{E}$.
2. A map $\iota^{n*} : D \to \mathrm{nspec}(\widehat{E})$, called the *numerical labeling map.*

2.4.1 Numerical Spectrum

Recall, that for the algebra E the dual, $\mathrm{spec}(E)$, consists of non-zero functionals $\mathcal{X} \in E^*$ which are multiplicative, that is, $\mathcal{X}(fg) = \mathcal{X}(f)\mathcal{X}(g)$ for all $f, g \in E$. Equivalently, $\mathrm{spec}(E)$ can be defined as the collection of non-zero linear functionals $\mathcal{X} \in E^*$ which fit in the following commutative diagram:



where m is induced from pointwise multiplication on $\mathbb{C}[D]$. Since $\widehat{E}$ is not closed under pointwise multiplication we cannot define it's characters in the usual algebraic sense. Instead, we consider an approximate multiplication map $m_{\widehat{E}} : \widehat{E} \otimes \widehat{E} \to \widehat{E}$ given by:

$$m_{\widehat{E}}(f \otimes g) = P_{\widehat{E}} m(f \otimes g),$$

for all $f, g \in \widehat{E}$, where $P_{\widehat{E}}$ denotes orthogonal projection onto $\widehat{E}$. We define a numerical character as a non-zero linear functional $\widehat{\mathcal{X}} \in \widehat{E}^*$ that "nearly fits"

in the following commutative diagram:

$$\begin{array}{ccc} \widehat{E} \otimes \widehat{E} & \xrightarrow{m_{\widehat{E}}} & \widehat{E} \\ & {\scriptstyle \widehat{\chi}\otimes\widehat{\chi}} \searrow \quad \swarrow {\scriptstyle \widehat{\chi}} & \\ & \mathbb{C} & \end{array} \tag{2.6}$$

To make the notion of "nearly fitting" precise we introduce an energy functional $\mathcal{E} : \widehat{E}^* \to \mathbb{R}$ that measures how well a linear functional fits in diagram (2.6):

$$\mathcal{E}(\widehat{\chi}) = \left|\left|\widehat{\chi} \otimes \widehat{\chi} - \widehat{\chi} m_{\widehat{E}}\right|\right|.$$

If $\widehat{E}$ is exactly E then the non-trivial zero solutions of $\mathcal{E}$ are exactly the characters. If $\widehat{E}$ is a perturbation of E then the local minimas of $\mathcal{E}$ are perturbations of the characters. By computing a large collection of minimas of $\mathcal{E}$ and applying k-means clustering we can compute $|C|$ minima centroids. We call these centroids the numerical spectrum of $\widehat{E}$ and denote them by $\mathrm{nspec}(\widehat{E})$.

2.4.2 Numerical labeling map

Recall, that for the algebra E the dual of ι is the map $\iota^* : D \to \mathrm{spec}(E)$ taking points $d \in D$ to the algebra character $\mathrm{ev}(d) \circ \iota$, where $\mathrm{ev}(d)$ is evaluation at d. For the vector space $\widehat{E}$ the canonical embedding $\iota : \widehat{E} \to \mathbb{C}[D]$ is still well defined, however we have no guarantee that the functional $\mathrm{ev}(d) \circ \iota$ is one of the numerical characters. Therefore, we define ι^{n*} as a map $\iota^{n*} : D \to \mathrm{nchar}(\widehat{\mathrm{E}})$ taking points $d \in D$ to the numerical character closest to $\mathrm{ev}(d) \circ \iota$:

$$\iota^{n*}(d) = \underset{\widehat{\chi} \in \mathrm{nspec}(\widehat{E})}{\mathrm{argmin}} \left|\left|\mathrm{ev}_d \circ \iota - \widehat{\chi}\right|\right|.$$

2.4.3 Comparison with spectral clustering

It is interesting to compare the categorical optimization solution with the spectral clustering solution. Both approaches begin with the same estimation step, taking the top $|C|$ eigenvectors of the empirical affinity kernel matrix. The span of these eigenvectors, $\widehat{E}$, gives an estimate of the clean vector space E. Both approaches next perform a "hard decision" using $\widehat{E}$ to assign a labeling to the data set D. However, the two approaches perform the hard decision in different ways. Spectral clustering assigns a labeling by applying k-means clustering to $\widehat{E}$ while categorical optimization assigns a labeling by computing a numerical spectrum for $\widehat{E}$, a notion associated with the algebraic structure of E.

In practice the two approaches exhibit similar performance. This is because both approaches are limited by the same first step, computing the estimate $\widehat{E}$ of E using the exact same eigenvector calculation. The Davis sin theorem shows that this eigenvector computation experiences a phase transition [29]. We illustrate this phase transition through the following numerical example. We set D to be $1,600$ data points containing two classes of size 800. The error rate of the empirical affinity kernels (probability of both false positive and false negatives) was increased from 0 to .5 in .05 increments. For each error rate, a random collection of empirical affinity kernels was generated, and the subspace $\widehat{E}$ was computed by taking the top two eigenvectors of the resulting empirical affinity kernel matrix. Figure 2.1 displays the relationship between the kernel error rate and the quality of the estimate $\widehat{E}$. Note the phase transition that occurs when the error rate approaches 0.45:

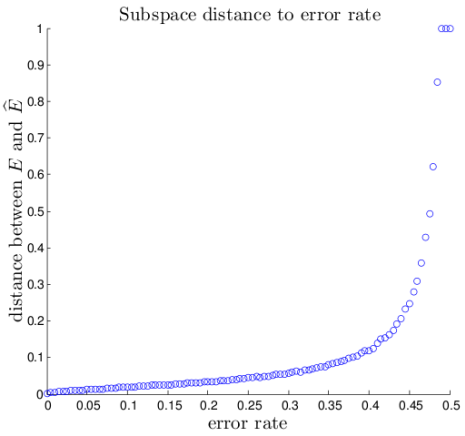


Figure 2.1: Illustration of phase transition in the eigenvector computation of $\widehat{E}$.

The phase transition of the eigenvector computation explains why the approaches exhibit similar performance. Before the phase transition, $\widehat{E}$ gives a very good estimate of E and so both approaches return an accurate labeling. After the phase transition, $\widehat{E}$ is a very poor estimate of E and so both approaches return an inaccurate labeling.

Chapter 3

Labeling geometrically parameterized data

We now consider the problem of labeling geometrically parameterized data. Our data set D is now equipped with both an unknown classification (C, c) and a known geometric parameterization (X, x). Also, we no longer have affinity kernels for each pair of data points. Instead, we have as input affinity kernels between all pairs of data points whose geometric parameters are within distance r of each other:

$$\{\delta_c(d, d') : m_X(x(d), x(d')) \leq r\}.$$

3.1 Local graph Laplacian

In this section we demonstrate how a (normalized) local graph Laplacian operator can be used to construct a solution to the labeling problem for geometrically parameterized data. Recall, that in section 2 we showed that the spectrum of the the graph Laplacian, $\mathcal{G}$, recovers the algebra E which, in turn, can be used to compute an admissible labeling of D. Since we no longer have access to the complete collection of affinity kernels, we cannot construct the graph Laplacian. Instead, we construct a local graph Laplacian operator, $\mathcal{G}_r : \mathbb{C}[D] \to \mathbb{C}[D]$, defined by:

$$(\mathcal{G}_r f)(d) = \frac{1}{|B_r(x(d))|} \sum_{d' \in N_r(d)} \delta_c(d, d') f(d'),$$

for all $f \in \mathbb{C}[D]$, where $B_r(x(d)) \subset X$ is the ball of radius r centered at $x(d)$ and $N_r(d)$ is the collection of all data points whose geometric parameter is within distance r of d's geometric parameter:

$$N_r(d) = \{d' : m_X(x(d), x(d')) \leq r\}.$$

The key point is that the spectrum of the local graph Laplacian operator, $\mathcal{G}_r$, can be expressed in terms of the spectrum of an operator $\mathcal{C}_r : \mathbb{C}[X] \to \mathbb{C}[X]$ of geometric origin. The operator is called the *cap operator* and is defined by:

$$(\mathcal{C}_r g)(x) = \frac{1}{|B_r(x)|} \sum_{y \in B_r(x)} g(y),$$

for all $g \in \mathbb{C}[X]$. The cap operator takes as input a function $g \in \mathbb{C}[X]$ and outputs a new function $\mathcal{C}_r g$ which at each point $x \in X$ is computed by averaging g over a cap of radius r centered at x. The precise relation between the spectrum of $\mathcal{C}_r$ and $\mathcal{G}_r$ is the content of theorem 3. To simplify the statement of the theorem we make a technical assumption, that the map $c \times x : D \to C \times X$, defined by:

$$(c \times x)(d) = (c(d), x(d)),$$

is a bijection. This means that every combination of class and geometric parameter occurs exactly once in the data set D. In practice, the map $c \times x$ is injective with dense image in $C \times X$. In this case the spectral relation of theorem 3 holds approximately. Furthermore, it can be shown using standard arguments from random matrix theory that as the size of D increases the approximation becomes tighter. Finally, in what follows, we abuse the tensor product notation by denoting the pointwise product of two subsets $A, B \subset \mathbb{C}[D]$ by $A \otimes B$.

Theorem 3. *(**Spectrum of local graph Laplacian**) Let $\lambda_1 > ... > \lambda_{|X|}$ be the eigenvalues of $\mathcal{C}_r$ with corresponding eigenspaces V_n for $n = 1, ..., |X|$. The operator $\mathcal{G}_r$ has eigenvalues λ_n with corresponding eigenspaces $E \otimes x^* V_n$.*

For the proof see Appendix B. Note that $\mathcal{C}_r$ is a Markov chain, moreover, in in many applications $\mathcal{C}_r$ is also ergodic (e.g. when X a large sampling of a path connected manifold). In such cases $\lambda_1 = 1$ with V_1 equal to the space of constant functions. In summary: if $\mathcal{C}_r$ is ergodic then it's top $|C|$ eigenvectors recover the algebra E, realizing the estimation stage of a solution to the labeling problem.

3.2 Biasdness of the local graph Laplacian

In practice we only have access to local empirical affinity kernels, $\widehat{\delta}_c(d, d') = \delta_c(d, d') + \epsilon(d, d')$, which define an empirical local graph Laplacian operator, $\widehat{\mathcal{G}}_r$. In this section we show that the noise term, $\epsilon(d, d')$, causes the spectrum of $\widehat{\mathcal{G}}_r$ to become biased to the geometry of X, thus causing k-means clustering to classify the data points by geometry rather then class membership. In what follows we use the noise model introduced at the end of section 2.3, namely, any random noise which gives an expected empirical affinity kernel of the form:

$$\mathbb{E}\widehat{\delta}_c(d, d') = (1 - q)\delta_c(d, d') + p(1 - \delta_c(d, d')),$$

for all d, d' with $m_X(x(d), x(d')) < r$, where $0 \leq p, q < 1$ are fixed constants. Under this assumption the expected operator takes the following form.

$$
\begin{aligned}
(\mathbb{E}\widehat{\mathcal{G}}_r f)(d) &= \frac{1}{|B_r(x(d))|} \sum_{d' \in N_r(d)} [(1-q)\delta_C(d,d') + p(1-\delta_C(d,d'))] f(d') \\
&= \frac{1}{|B_r(x(d))|} \sum_{d' \in N_r(d)} [(1-q-p)\delta_C(d,d') + p] f(d') \\
&= (1-q-p)\mathcal{G}_r f(d) + p\frac{1}{|B_r(x(d))|} \sum_{d' \in N_r(d)} f(d'),
\end{aligned} \tag{3.1}
$$

for all $f \in \mathbb{C}[D]$. Equation (3.1) expresses $\mathbb{E}\widehat{\mathcal{G}}_r$ as a sum of two operators. The first term is proportional to the clean local graph Laplacian, $\mathcal{G}_r$, reflecting both the geometry of X and the class membership relations. The second term is proportional to the cap operator and reflects only the geometry of X. As p increases the influence of the second term overwhelms the influence of the first term, causing $\mathbb{E}\widehat{\mathcal{G}}_r$ to essentially become a geometric object. The following theorem quantifies this phenomena.

Theorem 4. *(**Spectrum of expected local graph Laplacian**) Let $\lambda_1 > ... > \lambda_{|X|}$ be the eigenvalues of $\mathcal{C}_r$ with corresponding eigenspaces V_n for $n = 1, ..., |X|$. Let $\mathbb{C}[C]_0$ denote the space of constant functions on C and $\mathbb{C}[C]_0^{\perp}$ its orthogonal complement. Let E_0 and $E_0^{\perp}$ denote $c^*\mathbb{C}[C]_0$ and $c^*\mathbb{C}[C]_0^{\perp}$ respectively.*

1. *$E_0 \otimes x^*\mathbb{C}[X]$ and $E_0^{\perp} \otimes x^*\mathbb{C}[X]$ are invariant under $\mathbb{E}\widehat{\mathcal{G}}_r$.*
2. *The eigenvalues of $\mathbb{E}\widehat{\mathcal{G}}_r$ on $E_0 \otimes x^*\mathbb{C}[X]$ are $\lambda_n[|C|p + (1-q-p)]$ with corresponding eigenspaces $E_0 \otimes x^*V_n$.*
3. *The eigenvalues of $\mathbb{E}\widehat{\mathcal{G}}_r$ on $E_0^{\perp} \otimes x^*\mathbb{C}[X]$ are $\lambda_n(1-q-p)$ with corresponding eigenspaces $E_0^{\perp} \otimes x^*V_n$.*

For the proof see Appendix B. Theorem 4 shows that the presence of noise splits the $\mathcal{G}_r$ eigenspace $E \otimes x^* V_n$ into two pieces, $E_0 \otimes x^* V_n$ and $E_0^{\perp} \otimes x^* V_n$. The first piece is defined entirely in terms of the geometry of X and thus is uninformative for labeling, and the second piece is defined in terms of both the geometry of X and class membership. Thus the spectrum of $\mathbb{E}\widehat{\mathcal{G}}_r$ is an interlacing of eigenvalues corresponding to informative and non-informative eigenspaces (with the top eigenvalues corresponding to uninformative eigenspaces).

The following numerical experiment is an illustration of this point. A finite metric space S was generated by sampling 1000 points at random from the unit circle $x^2 + y^2 = 1$ lying in the xy plane of $\mathbb{R}^3$. The data set was taken to be the points in S together with their translation by $0.2e_3$. The data set consists of two parallel circles and is displayed in figure 3.1. The set-up is summarized in the following table:

Data Set	$D = S \cup (S + .2e_3)$
Parameterization	$(X, x) = (S, \pi_S)$
Classification	$(C, c) = (\{0, 0.2\}, e_3^*)$

Table 3.1: Parameterization and classification for the problem of labeling two parallel circles.

where $\pi_S : D \to S$ denotes projection onto S. We took r to be $10/(180\pi)$ giving 20° caps for neighborhoods. The error model parameters p and q were both taken to be 0.1. Figure 3.2 shows that the spectrum of $\mathbb{E}\widehat{\mathcal{G}}_r$ is an interlacing between eigenvalues corresponding to informative and uninformative eigenspaces. Figure 3.3 shows the result of applying k-means to

the top 15 eigenvectors of $\mathbb{E}\widehat{\mathcal{G}}_r$. As one can see, the data points are segmented by geometry rather then by class.

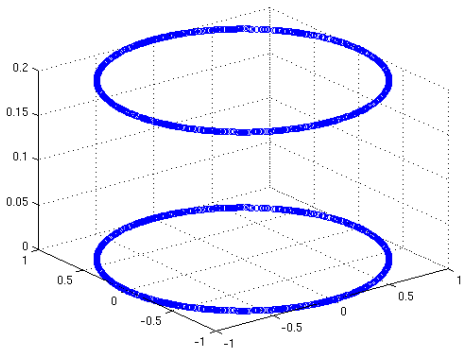


Figure 3.1: Data set D. The goal is to assign a labeling to D where all points with z coordinate 0 have a single label and all points with z coordinate 0.2 a different one.

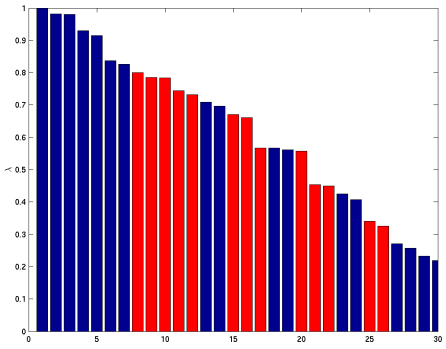
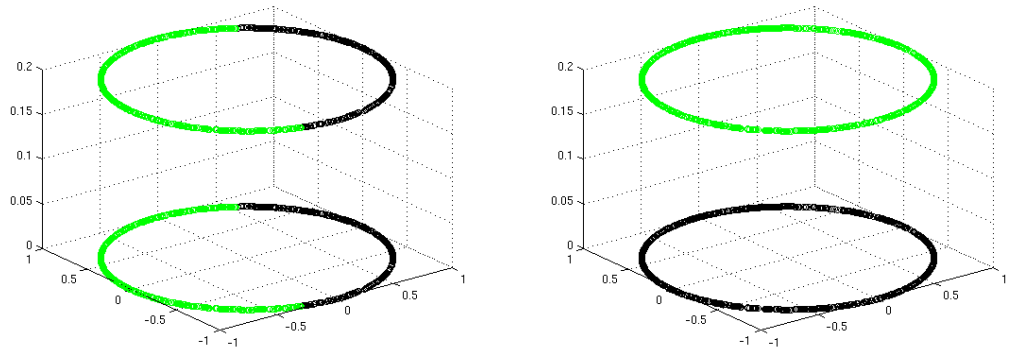


Figure 3.2: Top 30 eigenvalues of $\mathbb{E}\widehat{\mathcal{G}}_r$. Blue eigenvalues have corresponding eigenspaces contained in $E_0 \otimes x^*\mathbb{C}[X]$. Red eigenvalues have corresponding eigenspaces contained in $E_0^{\perp} \otimes x^*\mathbb{C}[X]$.



(a) Computed labeling of D (b) An admissible labeling of D

Figure 3.3: Result of applying k-means to the top 15 eigenvectors of $\mathbb{E}\widehat{\mathcal{G}}_r$. The data is segmented by geometry instead of by class membership.

Chapter 4

Local to Global Labeling of Parameterized Data

In this chapter we present a categorical optimization solution to the labeling problem for geometrically parameterized data, which is unbiased and robust to noise. The main idea is that since affinity kernels are known locally it is possible to solve the labeling problem for small patches of D in a robust manner. These local solutions can then be stitched together to give a global solution. With this in mind, we consider coverings $\mathcal{U} \subset \mathcal{P}(D)$ of D. Given a covering $\mathcal{U}$ we define an adjacency relation, $\mathrm{adj}_{\mathcal{U}}$, by:

$$\mathrm{adj}_{\mathcal{U}}(\mathrm{U}, \mathrm{W}) = \begin{cases} 1 & \text{if} \quad \mathrm{U} \cap \mathrm{W} \neq \emptyset \\ 0 & \text{if} \quad \mathrm{U} \cap \mathrm{W} = \emptyset \end{cases}$$

for all $U, W \in \mathcal{U}$. We call a covering $\mathcal{U}$ a *good covering* if the following two conditions are satisfied.

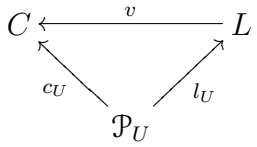
1. For every $U \in \mathcal{U}$ and for every $d, d' \in U$, we have $m_X(x(d), x(d')) \leq r$.
2. The graph induced by $\mathrm{adj}_{\mathcal{U}}$ is connected.

The first condition ensures that the labeling problem can be solved robustly on each set $U \in \mathcal{U}$. The second condition is necessary for these local solutions to be combined into a global solution. In practice good coverings are abundant. For example, if the parameter set X is a heavy sampling of a path connected manifold then the collection of neighborhoods, $N_{\frac{r}{2}}(d)$ for all $d \in D$, is a good covering.

4.1 Category of *Admissible partition labelings*

Suppose we are given a good covering $\mathcal{U}$. Then we can partition each $U \in \mathcal{U}$ by class. We denote the partition by $\mathcal{P}_U$. The classification map $c : U \to C$ is constant on partition sets, therefore, it defines a map $c_U : \mathcal{P}_U \to C$. The map c_U relates the local classification $\mathcal{P}_U$ to the global classification C and can be interpreted as a labeling of the partition set $\mathcal{P}_U$. This motivates the following definition.

Definition 6. *An **admissible partition labeling** is a pair $(L, \{l_U\}_{U\in\mathcal{U}})$ where L is a finite set and $l_U : \mathcal{P}_U \to L$ such that there exits a unique bijection (relabeling) $v : L \to C$ which fits in the following commutative diagram:*



(4.1)

for all $U \in \mathcal{U}$.

The collection of all admissible partition labelings forms a category where all objects are isomorphic to one another in a unique manner. Admissible partition labelings can be used to construct admissible labelings. Suppose we are given an admissible partition labeling $(L, \{l_U\}_{U\in\mathcal{U}})$. Let $p_U : U \to \mathcal{P}_U$ be the natural projection map which sends each data point to the unique partition set that contains it. The admissibility condition implies that the maps $l_U \circ p_U$ and $l_W \circ p_W$ agree on the intersection of their domains, $U \cap W$, for every $U, W \in \mathcal{U}$. Therefore, the collection of local labelings, $l_U \circ p_U$ for all $U \in \mathcal{U}$, defines a global labeling $l : D \to L$ which is admissible.

Invoking the duality between finite sets and finite algebras we replace the sets $C, \mathcal{P}_U$, and L with the algebras $\mathbb{C}[C], \mathbb{C}[\mathcal{P}_U]$, and $\mathbb{C}[L]$ and the maps ρ_U, c_U, v with the algebra morphisms ρ_U^*, c_U^*, v^*, respectively. In dual terms diagram (4.1) translates to the following commutative diagram:

$$\begin{array}{ccc} \mathbb{C}[C] & \xrightarrow{\quad v^* \quad} & \mathbb{C}[L] \\ & {\scriptstyle c_U^*} \searrow \quad \swarrow {\scriptstyle \rho_U^*} & \\ & \mathbb{C}[\mathcal{P}_U] & \end{array} \tag{4.2}$$

for all $U \in \mathcal{U}$. This motivates the following definition.

Definition 7. *An **admissible partition algebra** is a pair $(A, \{\alpha_U\}_{U\in\mathcal{U}})$ where A is an algebra and $\alpha_U : A \to \mathbb{C}[\mathcal{P}_U]$ is an algebra morphism such that there exits a unique algebra isomorphism $\gamma : A \to \mathbb{C}[C]$ which fits in the following commutative diagram:*

$$\begin{array}{ccc} \mathbb{C}[C] & \xrightarrow{\quad \gamma \quad} & A \\ & {\scriptstyle c_U^*} \searrow \quad \swarrow {\scriptstyle \alpha_U} & \\ & \mathbb{C}[\mathcal{P}_U] & \end{array} \tag{4.3}$$

for all $U \in \mathcal{U}$

Given an admissible partition algebra $(A, \{\alpha_U\}_{U\in\mathcal{U}})$, the dual pair $(\text{spec}(A), \{\alpha_U^*\}_{U\in\mathcal{U}})$ is an admissible partition labeling. To see this, take the dual of diagram (4.3):

$$\begin{array}{ccc} C & \xleftarrow{\quad \gamma^* \quad} & \text{spec}(A) \\ & {\scriptstyle c_U} \nwarrow \quad \nearrow {\scriptstyle \alpha_U^*} & \\ & \mathcal{P}_U & \end{array} \tag{4.4}$$

for all $U \in \mathcal{U}$. Because diagram (4.3) commutes, the functorality of pullbacks implies that this diagram commutes as well. Therefore, we conclude that $(\text{spec}(A), \{\alpha_U^*\}_{U\in\mathcal{U}})$ is an admissible partition labeling.

4.2 Construction of an admissible partition algebra

In this section we explain how to construct a particular admissible partition algebra. We follow the same approach used in section 2.2. We first embed $\mathbb{C}[C]$ as a subalgebra F of a high dimensional algebra H. We next identify an operator on H whose spectrum recovers the subalgebra $F \subset H$. Finally, we show how clean (that is error-free) local partitions can be used to construct the operator.

4.2.1 Embedding the algebra $\mathbb{C}[C]$

A high dimensional algebra H is defined as the global sections of an algebra bundle B, where:

$$B := \bigsqcup_{U \in \mathcal{U}} \mathbb{C}[\mathcal{P}_U] \to \mathcal{U}.$$

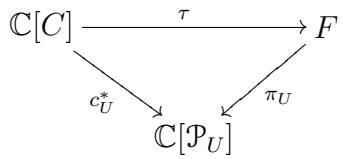
The algebra $\mathbb{C}[C]$ is embedded into H through the algebra morphism $\tau : \mathbb{C}[C] \to H$ given by:

$$(\tau f)(U) = c_U^*(f),$$

for all $f \in \mathbb{C}[C]$ and $U \in \mathcal{U}$. We denote the image of τ by F. The algebra F is equipped with a collection of canonical projection maps, $\pi_U : F \to \mathbb{C}[\mathcal{P}_U]$, given by:

$$\pi_U(s) = s(U),$$

for all $s \in F$ and $U \in \mathcal{U}$. Evidently the pair $(F, \{\pi_U\}_{U \in \mathcal{U}})$ is an admissible partition algebra, since the following diagram commutes:



for all $U \in \mathcal{U}$. Note that once the algebra F is computed, the π_U maps are given for free. Therefore, the problem of combining local labelings into a global labeling can be reduced to computing the algebra F.

4.2.2 Spectral characterization of F

The algebra F admits a nice geometric interpretation. Viewing the bijections $c_U : \mathcal{P}_{\mathcal{U}} \to C$ as a choice of coordinates on C, one defines transition maps, $t(W, U) : \mathcal{P}_U \to \mathcal{P}_W$, given by:

$$t(W, U) = (c_W)^{-1} \circ c_U,$$

for all pairs $U, W \in \mathcal{U}$. The subalgebra F consists of sections invariant under change of coordinates. This fact is the content the following proposition.

Proposition 1. ***(Geometric characterization of** F**)*** *A section s is an element of F if and only if the following condition holds:*

$$t(W, U)^* s(W) = s(U), \tag{4.5}$$

for all $U, W \in \mathcal{U}$.

For the proof see Appendix B. The statement of proposition 1 can be expressed as a spectral property. Let $\mathcal{T} : H \to H$ be an operator given by:

$$(\mathcal{T}s)(U) = \frac{1}{|N(U)|} \sum_{W \in N(U)} t(W, U)^* s(W), \tag{4.6}$$

for all $s \in H$, where $N(U) = \{W \in \mathcal{U} : U \cap W \neq \emptyset\}$. We call $\mathcal{T}$ the *transition operator*. The invariance condition (4.5) implies that sections in F are fixed by the transition operator, that is, $\mathcal{T}s = s$ for all $s \in F$. In other words the algebra F is contained in the eigenvalue one eigenspace of the transition operator.

To fully describe the spectrum of the transition operator we introduce a cap operator, $\mathcal{C} : \mathbb{C}[\mathcal{U}] \to \mathbb{C}[\mathcal{U}]$, given by:

$$(\mathcal{C}f)(U) = \frac{1}{|N(U)|} \sum_{W \in N(U)} f(W),$$

for all $f \in \mathbb{C}[\mathcal{U}]$. The relationship between the spectrum of $\mathcal{C}$ and $\mathcal{T}$ is the content of the following theorem.

Theorem 5. ***(Spectrum of the transition operator)*** *Let $\lambda_1 > ... > \lambda_{|\mathcal{U}|}$ be the eigenvalues of $\mathcal{C}$ with corresponding eigenspaces V_n for $n = 1, ..., |\mathcal{U}|$, then:*

1. *$\lambda_1 = 1$ with V_1 the space of constant functions.*
2. *The eigenvalues of $\mathcal{T}$ are λ_n with corresponding eigenspaces $F \otimes V_n$* [1]*.*

For the proof see Appendix B. Theorem 5 states that the top eigenspace of the transition operator recovers the algebra F.

4.2.3 Computing $\mathcal{T}$

We conclude this section with an explanation of how to compute transition maps from the clean local classifications. For each pair of overlapping covering sets $U, W \in \mathcal{U}$ do the following:

1. For each partition set $A \in \mathcal{P}_U$ take the intersection with W, giving the non-empty set $A \cap W$.

[1] We abuse tensor notation to denote for a pair of sets, $A \subset H$ and $B \subset \mathbb{C}[\mathcal{U}]$, their product as $A \otimes B$. Where multiplication is defined by: $(s \cdot f)(U) = s(U)f(U)$, for all $s \in H$ and $f \in \mathbb{C}[\mathcal{U}]$.

2. Map A to the partition set of $\mathcal{P}_W$ which contains $A \cap W$ as a subset, this will be exactly $t(W, U)A$.

From the computed transition maps, the transition operator is constructed using equation (4.6).

4.3 Effect of local misclassifications

The previous section showed how the clean local classifications can be used to construct the transition operator. In reality, the local classifications are not clean, that is, some points are misclassified. In this section we suggest a construction of an *empirical* transition operator, denoted $\widehat{\mathcal{T}}$, from non-perfect local classifications. We then show that misclassifications cause $\widehat{\mathcal{T}}$ to become biased to geometry. Finally, we show that by restricting the domain of $\widehat{\mathcal{T}}$ to a subspace of H, it becomes unbiased.

4.3.1 Empirical transition operator

The empirical transition operator is defined in terms of empirical transition maps, $\widehat{t}(W, U) : \mathcal{P}_U \to \mathcal{P}_W$, for overlapping covering sets $U, W \in \mathcal{U}$. Where $\widehat{t}(W, U)$ is defined as follows:

1. For each set $A \in \mathcal{P}_U$ take the intersection with W, giving the non-empty set $A \cap W$.

2. Define $\widehat{t}(W, U)(A)$ to be the partition set in $\mathcal{P}_W$ containing the most points of $A \cap W$.

Combining these maps together defines an empirical transition operator, denoted $\widehat{\mathcal{T}}$:

$$(\widehat{\mathcal{T}}s)(U) = \frac{1}{|N(U)|} \sum_{W \in N(U)} \widehat{t}(W,U)^* s(W),$$

for all $s \in H$ and $U \in \mathcal{U}$. We note that in the case of perfect local classifications the empirical transition operator $\widehat{\mathcal{T}}$ is exactly the clean transition operator $\mathcal{T}$.

4.3.2 Biasdness of $\widehat{\mathcal{T}}$

Errors in the local classifications cause errors in the empirical transition maps. We first consider a simple random error model for the empirical transition maps. Namely, with probability p the clean transition map is computed, and with probability $(1-p)$ a random bijection is computed. This error model yields the following stochastic kernel:

$$\widehat{t}(W,U) = \begin{cases} t(W,U) \text{ with probability } p \\ \text{a random element of } S(\mathcal{P}_W, \mathcal{P}_U) \text{ with probability } 1-p \end{cases}$$

where $S(\mathcal{P}_W, \mathcal{P}_U)$ denotes the collection of all bijections from $\mathcal{P}_U$ to $\mathcal{P}_W$. The expected value of the pullback under this error model is given by:

$$\mathbb{E}[\widehat{t}(W,U)^*] = pt(W,U)^* + (1-p)P_0(U,W), \tag{4.7}$$

where:

$$P_0(U,W) = \frac{1}{|C|!} \sum_{b \in S(\mathcal{P}_W, \mathcal{P}_U)} b^*.$$

The operator $P_0(U,W)$ projects onto the space of constant functions, to see this note that:

$$(P_0(U,W)f)(A) = \frac{1}{|C|} \sum_{B \in \mathcal{P}_W} f(B),$$

for all $f \in \mathbb{C}[\mathcal{P}_W]$ and $A \in \mathcal{P}_U$. Using equation (4.7) we can express the expected empirical tranisition operator as a sum of two terms:

$$\mathbb{E}\widehat{\mathcal{T}} = p\mathcal{T} + (1-p)P_0. \tag{4.8}$$

Where the operator $P_0 : H \to H$ projects pointwise onto the one dimensional space of constant functions:

$$(P_0 s)(U) = P_0(U,U)s(U),$$

for all $s \in H$ and $U \in \mathcal{U}$. Note that only the first term in (4.8) carries information about classification. Therefore, the noise causes the eigenspaces of $\mathcal{T}$ to split into two pieces, one defined solely in terms of the adjacency of $\mathcal{U}$ and the second defined by a combination of adjacency and class. This phenomena is quantified in the following theorem.

Theorem 6. ***(Spectrum of the expected transition operator)*** *Let $\lambda_1 > ... > \lambda_{|\mathcal{U}|}$ be the eigenvalues of $\mathcal{C}$ with corresponding eigenspaces V_n for $n = 1, ..., |\mathcal{U}|$. Let $\mathbb{C}[C]_0$ denote the constant function on C and $\mathbb{C}[C]_0^{\perp}$ its orthogonal complement. Let F_0 and $F_0^{\perp}$ denote $\tau\mathbb{C}[C]_0$ and $\tau\mathbb{C}[C]_0^{\perp}$ respectively, then:*

1. *$F_0 \otimes \mathbb{C}[\mathcal{U}]$ and $F_0^{\perp} \otimes \mathbb{C}[\mathcal{U}]$ are invariant under $\mathbb{E}\widehat{\mathcal{T}}$.*
2. *The eigenvalues of $\mathbb{E}\widehat{\mathcal{T}}$ on $F_0 \otimes \mathbb{C}[\mathcal{U}]$ are λ_n with corresponding eigenspaces $F_0 \otimes V_n$.*
3. *The eigenvalues of $\mathbb{E}\widehat{\mathcal{T}}$ on $F_0^{\perp} \otimes \mathbb{C}[\mathcal{U}]$ are $p\lambda_n$ with corresponding eigenspaces $F_0^{\perp} \otimes V_n$.*

4.3.3 Restricting $\widehat{\mathcal{T}}$

Theorem 6 states that the spectrum of $\mathbb{E}\widehat{\mathcal{T}}$ is an interlacing of eignvalues corresponding to informative and uninformative eigenspaces (from the perspective of classification). The main point is we know the non-informative subspace $F_0 \otimes \mathbb{C}[\mathcal{U}]$ is contained in the subspace of sections which are fiberwise constant, we denote this space by H_0:

$$H_0 = \Gamma\left(\bigsqcup_{U \in \mathcal{U}} \mathbb{C}[\mathcal{P}_U]_0\right).$$

Likewise, we know that the informative subspace $F_0^{\perp} \otimes \mathbb{C}[\mathcal{U}]$ is contained in the subspace of sections which are fiberwise orthogonal to constant functions, we denote this subspace by $H_0^{\perp}$:

$$H_0^{\perp} = \Gamma\left(\bigsqcup_{U \in \mathcal{U}} \mathbb{C}[\mathcal{P}_U]_0^{\perp}\right).$$

By restricting $\mathbb{E}\widehat{\mathcal{T}}$ to $H_0^{\perp}$, the resulting spectrum will contain only informative eigenspaces. In particular, the top eigenspace recovers $F_0^{\perp}$. Furthermore, we note that F_0 is the one dimensional space of constant sections, and so is known a-priori. Therefore, once $F_0^{\perp}$ is computed one gets $F = F_0 \bigoplus F_0^{\perp}$ for free.

To summarize: under the error model of random bijections by restricting the operator $\widehat{\mathcal{T}}$ to $H_0^{\perp}$, the operator become unbiased to noise. Note that we can generalize the error model of the transition maps to any which have an expected pullback given by:

$$\mathbb{E}[\widehat{t}(W,U)^*] = pt(W,U)^* + qP_0(U,W),$$

where p and q are fixed constants. This includes more realistic error models, for example, empirical tranisition maps which return the true transition map up to two random errors.

4.4 Numerical Dualization

Given non-clean local classifications of the covering sets in $\mathcal{U}$, we can use the procedure of the previous section to define an empirical transition operator, $\widehat{\mathcal{T}}$. Restricting it to $H_0^{\perp}$ and taking the span of it's top $|C|-1$ eigenvectors defines a subspace which we denote by $\mathbb{V}$. Let $\widehat{F}$ denote $\mathbb{V} \bigoplus 1$, then as a vector space $\widehat{F}$ is close to the algebra F (this follows from theorem 6 and standard results from random matrix theory). In this subsection we will show how $\widehat{F}$ together with the non-perfect local labelings can be used to robustly compute a global labeling of D.

Although $\widehat{F}$ is close to F as a vector space, there is no guarantee that $\widehat{F}$ is an algebra. Therefore, we cannot invoke standard duality to transform the pair $(\widehat{F}, \{\pi_U\}_{U \in \mathcal{U}})$ into an admissible partition labeling. Instead, we use the numerical dualization method introduced in section 2.4 to compute a numerical spectrum, $\text{nspec}(\widehat{F})$, and numerical labeling maps, $\pi_U^{n*} : \mathcal{P}_U \to \text{nspec}(\widehat{F})$, thus giving the pair:

$$\left(\text{nspec}(\widehat{F}), \{\pi_U^{n*}\}_{U \in \mathcal{U}}\right).$$

The pair will not be an admissible partition labeling, therefore, we can't translate it to a global admissible labeling using algebra. Instead, we use voting to choose a global labeling for D. For each data point $d \in D$ do the following:

1. Find all the neighborhoods $U \in \mathcal{U}$ containing d.
2. For each such neighborhood U see what label d gets sent to under the local labeling $\pi_U^{n*} \circ p_U$.
3. Set the global label of d to be the label d is sent to most under the local labelings.

If all the local labelings of the data points are error-free, then this procedure constructs a global admissible labeling. If the local labelings of the data points have errors, then the procedure suppresses them. To understand this we interpret the voting procedure using naive Bayes.

4.4.1 Naive Bayes interpretation

For a data point $d \in D$ let the collection of partition sets containing d be denoted by $\mathcal{U}_d$. Then the local labels of d can be interpreted as features:

$$f_U := \pi_U^{n*} \circ p_U(d),$$

for all $U \in \mathcal{U}_d$. Suppose we take the conditional probabilities of the features (with respect to the labels) to be:

$$\Pr(f_U|l) = \begin{cases} 1-\epsilon & \text{if} \quad f_U = l \\ \epsilon & \text{if} \quad f_U \neq l \end{cases}$$

for all $l \in \text{nspec}(\widehat{F})$, where ϵ is some constant with $0 < \epsilon < 1/2$. Then assuming a uniform prior for the different labels, naive Bayes gives:

$$\Pr(l) = \frac{1}{|C|}(1-\epsilon)^n \epsilon^{(|\mathcal{U}_d|-n)},$$

where n denotes the number of sets in $\mathcal{U}_d$ with $f_U = l$. Under this model, the most likely label for the data point d is given by the voting procedure of section 4.4. The effectiveness of the voting procedure can now be understood, as long as the majority of local labelings are correct, the voting procedure correctly labels the point. Note that the naive Bayes interpretation suggests the voting procedure can be improved by more accurate conditional probabilities and priors for the labels. One way to do this is through expectation maximization between these unknown parameters and the unknown global labeling of the data.

Chapter 5

Numerical Results

The categorical optimization solution of chapter 4 was applied to the heterogeneous Cryo-EM problem. A phantom was constructed in the following way. Fifty balls were randomly chosen by selecting for each ball a center randomly using the uniform measure on the cube $[-5,5] \times [-5,5] \times [-5,5]$ and a radius randomly using the uniform measure on the unit interval. The union of these fifty balls gave the phantom. A second phantom was generated by randomly removing one of these balls. Figure 5.1 shows the two phantoms. The ratio of the volume of the removed ball to the total volume of the first phantom was 5/100. The phantoms were used to construct potential functions which took the value one on the union of the balls and zero elsewhere. An electron microscope simulator [20] was used to generate $2,500$ projection images from random orientations of the first phantom and $2,500$ projection images from random orientations of the second phantom. White noise was added to the images to give an snr of 1/30. The noisy projection images and their orientations were taken as input, the desired output was a labeling of the data consistent with phantom type, i.e. two images have the same label if and only if they were generated from the same phantom.

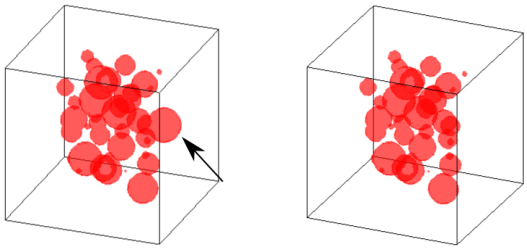


Figure 5.1: The two generated phantoms. The presence of a single ball distinguishes the left phantom from the right phantom.

We proceed to translate this problem into the categorical framework introduced in this thesus; namely we identify a data set, a geometric parameterization, and a classification. The data set is the observed collection of $5,000$ projection images, denoted by I. The geometric parameter set consists of the $5,000$ orientations, denoted by O. The orientations are represented as matrices in the rotation group. We equip O with a pseudometric $m_O : O \times O \to \mathbb{R}$ which returns a distance of zero between two images which differ only by a planer rotation, namely:

$$m_O(x, y) = d_{S^2}(x(e_3), y(e_3)),$$

for all pairs $x, y \in O$, where $d_{S^2} : S^2 \times S^2 \to \mathbb{R}$ is the standard metric on the unit sphere. The geometric parameterization is the map $o : I \to O$ taking images to their corresponding orientation. Finally, the intrinsic classification is the pair $(\{P_1, P_2\}, p)$. Where P_1 denotes the first phantom, P_2 denotes the second phantom, and p is a map $p : I \to \{P_1, P_2\}$ which assigns each image to the phantom used to generate it.

Data Set	$D = I$
Parameterization	$(X, x) = (O, o)$
Classification	$(C, c) = (\{P_1, P_2\}, p)$

Table 5.1: Parameterization and classification for the heterogeneous labeling problem.

We now show how the categorical optimization solution presented in chapter 4 was used to construct a labeling of I. As a prepossessing step we applied rotationally invariant principal component analysis filtering [36] to increase the snr of the images. Figure 5.2 shows the effect of the PCA filtering.

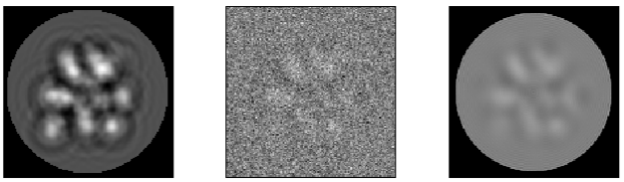


Figure 5.2: Left: clean projection image. Middle: result of white noise added to left image. Right: result of PCA filtering applied to middle image.

Next, we constructed an admissible covering of I. This was done by selecting $1,000$ images from I at random, the collection of balls (defined by the pseudometric m_O) centered at these sampled images and having radius $10\pi/180$ gave an admissible covering, denoted by $\mathcal{U}$. Next, for each covering set $U \in \mathcal{U}$ the images within U were rotationally aligned and the resulting images were partitioned by class using sparse subspace clustering [3]. We denote the resulting partition by $\mathcal{P}_U$. These partitions were used to construct an empirical transition operator, $\widehat{\mathcal{T}}$, using the procedure presented in section 4.2. To eliminate the bias of noise the operator was restricted to the subspace $H_0^{\perp}$. The top eigenspace, $\mathbb{V}$, of the restricted operator was computed. The

space $\mathbb{V}$ was combined with the trivial space 1 to give the algebra estimate, $\widehat{F}$. The numerical dualization procedure described in chapter 2.4 was used to compute a numerical spectrum for $\widehat{F}$. Finally the voting procedure described in section 4.4 used the local partitions and the numerical spectrum to compute a labeling of I. The resulting labeling was 98.6% accurate.

Chapter 6

Conclusion

6.1 Summary

In the first part of the thesis we considered the labeling problem for geometrically paramerized data. We gave three examples of this problem, and showed that the central difficulty of the problem is that data point variation due to difference in geometric parameter often overwhelms data point variation due to difference in class. We showed that this fact causes the spectral clustering method to become biased to noise, and quantified the bias in theorem 4. The main result was a categorical optimization solution to this labeling problem. The solution works by first solving the labeling problem over small patches of data points and then combining theses local labelings into a global labeling. The robustness of the solution was established by theorem 6.

6.2 Future work

6.2.1 Categorical optimization for general classification

The categorical framework developed in this thesis can be applied to more general classification problems. For example, consider the problem of identifying the party affiliation of U.S. congressmen from voting records [21]. For each bill we assume that the Republican and Democratic party takes a

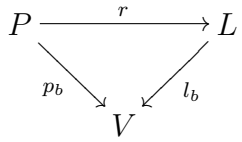
unique position. This can be expressed as a bijection p_b with:

$$p_b : P \to V,$$

where b denotes a certain bill, $V = \{\text{Yes}, \text{No}\}$, and $P = \{\text{Republican}, \text{Democrat}\}$. Once the maps p_b are known, naive Bayes can be used to assign a party affiliations to the congressmen based on their voting records. To this end we note that the transition maps:

$$p_{b'} \circ p_b^{-1} : V \to V,$$

can be robustly computed from the voting records of bills b and b'. Using these transition maps and categorical methods analogous to those introduced in this thesis the maps p_b can be recovered up to a relabeling. That is, we can compute a two element set L and a collection of maps $l_b : L \to V$ such that there exits a unique bijection (relabeling) $r : P \to L$ fitting in the following diagram:



for all bills b. The maps l_b can then be used in conjunction with naive Bayes to assign affiliations to the congressmen.

6.2.2 Categorical optimization for probabilistic data structures

Categorical optimization currently can only be applied to problems having rigid algebraic structures. However, most machine learning problems have probabilistic rather then algebraic structure. For example, in the problem of congressional voting the maps $p_b : P \to V$ would be more accurately modeled as maps:

$$p_b : P \to \Pr(V),$$

that is, each party has certain probability of voting yes or no for bill b. I believe an interesting question of research would be: can categorical optimization be extended to data with probabilistic structure?

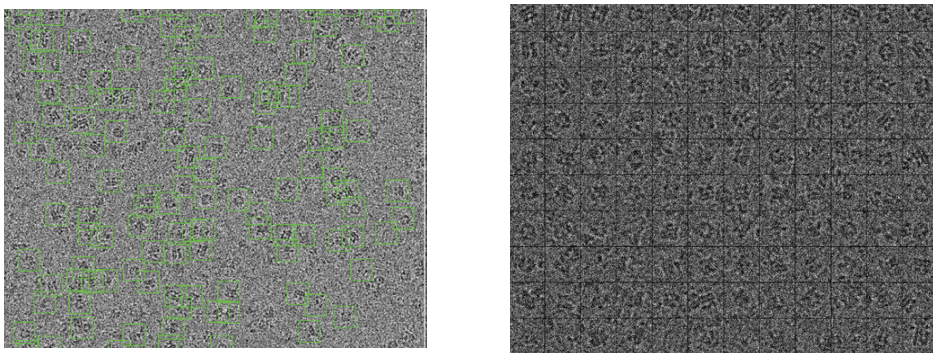
Part II

Denoising and centering Cryo-EM images by global projection

Chapter 7

Introduction

Single particle Cryo-Electron Microscopy (Cryo-EM) is an imaging technique used to find the potential function of biomolecules. It works by freezing a large collection of randomly oriented copies of a single biomolecule (on the order of 10,000) in vitreous ice, and imaging the sample with an electron beam. The 2D projection images of the biomolecules are extracted (boxed out) of the resulting micrograph, thus producing a stack of images. Using this image stack a 3D model of the potential function can be constructed [6].

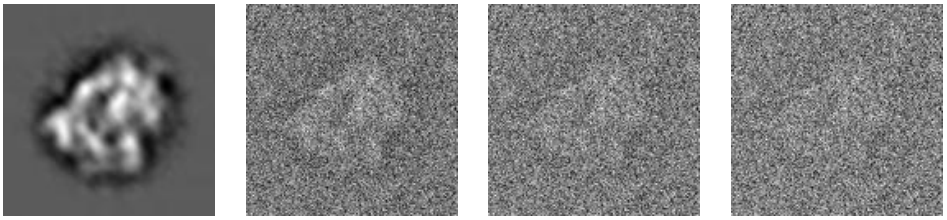


(a) Micrograph of sample. (b) Stack of boxed particles.

Figure 7.1: Projection and particle stack of a TRPV1 sample. Figure taken from "Structure of the TRPV1 ion channel determined by electron cryomicroscopy" [16].

The primary challenge of Cryo-EM is that the images have extremely low SNR, this is because to avoid radiation damage and sample movement the

biomolecules can only be exposed to the electron beam for a short time [9]. We now review how practitioners deal with the problem of low image SNR.



(a) Clean image. (b) SNR $\frac{1}{20}$ (c) SNR $\frac{1}{50}$ (d) SNR $\frac{1}{100}$

Figure 7.2: Simulated projection image with increasing amount of added Gaussian noise.

7.1 Denoising methods

The three methods used in practice to increase the SNR of Cryo-EM images are: low-pass filtering, principal component analysis filtering, and class averaging. When evaluating the methods we keep in mind three important questions: How well does the method remove noise? How well does the method preserve signal? Does the method improve with increasing number of input images?

- **Low-pass filtering:** Noisy images with p by p pixels are viewed as vectors in a p^2 dimensional vector space. A low dimensional vector space V_{filter} is identified which well contains the energy of the signal (the true images). The low dimensional space is typically the span of low-frequency functions. The noisy images are denosised by projection onto V_{filter}, all noise (and signal) perpendicular to V_{filter} is removed. There is a trade-off between signal preservation and noise reduction, choosing a larger V_{filter} preserves more signal but removes less noise.

Because the filtering is performed independently for each image, low-pass filtering does not improve with increasing number of images. In practice Wiener filters (that is a weighted projection) are used instead of a straight low-pass filter [35], [19], [23].

- **Principal component analysis filtering:** Fix a potential function ϕ, assuming image noise is isotropic (not biased to any direction), the optimal low-pass filter of dimension k would be the k dimensional subspace that best contains the energy of the true projection images of ϕ. That is, $V_{k,opt}$ is the solution to the following optimization problem:

$$\begin{aligned} \underset{V}{\text{minimize}} \quad & \mathbb{E}\|P_V I\|^2 \\ \text{subject to} \quad & \dim(V) = k, \end{aligned}$$

where I is a projection image of ϕ taken from a random orientation, P_V denotes orthogonal projection onto V, and expectation is taken with respect to the random orientations. Principal component analysis (PCA) takes as input a collection of noisy images and computes the k dimensional subspace which best contains the energy of the noisy images, we denote the subspace by $\widehat{V}_{k,opt}$. It can be shown that as the number of noisy images goes to infinity $\widehat{V}_{k,opt}$ approaches $V_{k,opt}$. The images are then denoised by using $\widehat{V}_{k,opt}$ as a low pass filter. The performance of PCA filtering asymptotes with increasing number of images: at first increasing the number of images improves the PCA estimate of $V_{k,opt}$, however, even with a very good estimate of $V_{k,opt}$ PCA filtering has the same limitation as any low-pass filter: only noise orthogonal to $\widehat{V}_{k,opt}$ is removed, any noise in $\widehat{V}_{k,opt}$ stays. This is true no matter how many images are used. For a good review of PCA in Cryo-EM see [27]. For the state of the art PCA algorithm see [36].

- **Class Averaging:** Classes of images are identified, where each class contains images similar up to translation, rotation, and noise. Images within each class are then aligned and averaged to produce "class averages" having higher SNR. The size of each class grows linearly with the number of images. Since images within a class are aligned and averaged this means the SNR of the class averages grows linearly with the number of images. However, the amount of signal loss is independent of the number of images. This is because images within a single class are only approximately the same, so when the images from a class are aligned and averaged the resulting class average will experience blurring due to the inner-class discrepancy. The state of the art class averaging algorithm is given in [37].

In practice images are denoised using a combination of the methods. First, PCA filtering is used to compress and denoise the images. Next, the filtered images are used to identify classes of similar images. Finally, the original images within each class are aligned and averaged, producing higher SNR class averages.

7.2 Main results

The second half of the thesis introduces a novel algorithm, called *global projection*, which takes as input Cryo-EM image collections and outputs a denoised image collection. The algorithm has three important properties:

- The SNR of the output images grows linearly with the number of input images.

- The output images are centered.

The idea of the algorithm is simple, suppose we have a collection of n Cryo-EM images each with $p \times p$ pixels. Then the collection of images can be viewed as a single vector in a np^2 dimensional space (typical values are $n = 10^4$ and $p = 10^2$, giving a vector space of dimension 10^8). Suppose a low dimensional subspace can be identified within the np^2 dimensional vector space which well contains the signal (i.e. the vector of clean images), then projecting the vector of observed images onto this subspace gives an effective denoising algorithm. The technical portion of this part of the thesis will be identifying such a vector space and showing how it can be computed in practice.

7.3 Structure of part II

The second part of the thesis consists of six chapters and two appendices.

- **Chapter 8:** In this chapter we present the central geometric idea of the second part of the thesis: that collections of images from a Cryo-EM experiment can be interpreted as a single geometric object (a vector), and that the X-ray transform can be interpreted as giving a correspondence between potential functions and image collections. We formalize the geometric structure and correspondence using the language of vector bundles and sections. Next, we identify a finite dimensional vector space, denoted V_b, which well contains potential functions up to the band limit b. The subspace is defined using the Fourier-Bessel functions. Finally, we introduce the global projection algorithm. The algorithm works by

projecting the vector of observed Cryo-EM images onto a low dimensional subspace, denoted W_b (computing the space W_b is non-trivial), which is related to V_b by the X-ray transform.

- **Chapter 9:** In this chapter we present a method to compute W_b. We do this by first characterizing V_b and W_b in terms of the representation theory of $SO(3)$, and then showing how the X-ray transform relates the two vector spaces. Next, we demonstrate how image collections can be used to construct the *class averaging operator*, and how the spectrum of this operator can be used to recover the vector space W_b.

- **Chapter 10:** In this chapter we study what happens when the input images are not centered. We show that even when the input images are non-centered, global projection still returns as output a collection of centered images.

- **Chapter 11:** In this chapter we express the global projection algorithm in coordinates. Then, using this expression we present a detailed outline of an implementation of the global projection algorithm.

- **Chapter 12:** In this chapter we present the results from applying the global projection algorithm to simulated and real image collections.

- **Chapter 13:** In this chapter we give a summary and conclusion for the second part of the thesis. We then outline how the global projection algorithm could possibly be used to align direct detector movie frames.

- **Appendix C:** Contains proofs of propositions and theorems found in the second part of the thesis.

- **Appendix D:** Contains an overview of the charge transfer function, phase flipping, and defocus groups.

Chapter 8

Global projection

8.1 Geometric Structure of Image Collections

In this section we present the central geometric idea of the thesis: that collections of images from a Cryo-EM experiment can be interpreted as a single geometric object, and that the X-ray transform can be interpreted as giving a correspondence between potential functions and image collections.

In this thesis we view Cryo-EM images as coming from a single fixed potential function, with the microscope moving about the potential function taking projections from random orientations. This is equivalent to the actual Cryo-EM setup, where the images are projections of randomly oriented copies of a single potential function.

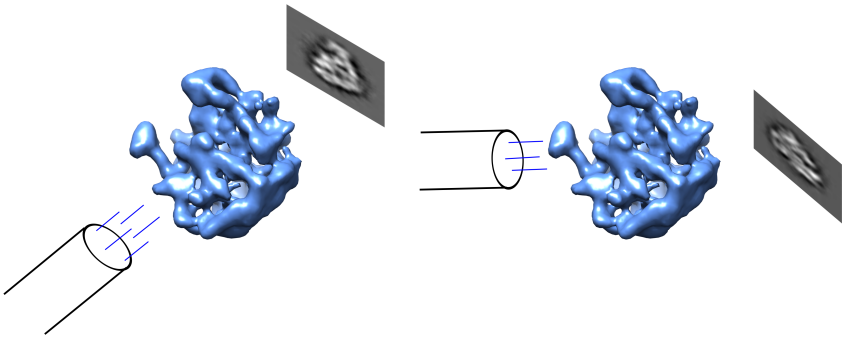


Figure 8.1: Potential function fixed, with the microscope moving about taking projections.

The following table fixes notation for the thesis. The first column contains different Cryo-EM concepts, the second column contains the mathematical objects used to model the concepts, finally the third column gives notation to the objects.

Cryo-EM Concept	**Mathematical Object**	**Notation**
Domain of potential functions	3D vector space equipped with an orientation and inner product.	V
Potential functions	Continuous functions compactly supported in the unit ball.	$C_0(B_V)$
Images	Continuous functions compactly supported in the $\mathbb{R}^2$ unit disk.	$\mathcal{I}$
Microscope Orientations	Orientation and inner product preserving embeddings of $\mathbb{R}^3$ into V.	$\mathcal{F}$

Table 8.1: Notation for Cryo-EM concepts.

Remark: The unit length of V and $\mathbb{R}^2$ depend on the Cryo-EM experiment. Suppose we are given a collection of projection images, following standard Cryo-EM procedure we apply a spherical mask to each projection image. The length of the mask radius, denoted r, depends on the size of the biomolecule being imaged (typically r is around $100 - 300$ angstroms). We set the unit length of both V and $\mathbb{R}^2$ to be r.

Up to first order Cryo-EM images are X-ray transforms of potential functions. The X-ray transform, denoted by X, takes as input a potential function and a microscope orientation and outputs an image through an integration

formula:

$$[X(\phi, x)](a, b) = \int_t \phi(ax(e_1) + bx(e_2) + tx(e_3))\,\mathrm{d}t,$$

for all $\phi \in C_0(B_V)$ and $x \in \mathcal{F}$. It will prove useful to express X in terms of the standard X-ray transform, denoted by X_{std}, which projects functions in $\mathbb{R}^3$ onto the standard xy plane:

$$(X_{std} f)(a, b) = \int_t f(a, b, t)\,\mathrm{d}t,$$

for all $f \in C_0(B_{\mathbb{R}^3})$. The X-ray transform applied to a potential function ϕ and a microscope orientation x can be understood in two steps: first it writes ϕ as a function in $\mathbb{R}^3$ using the microscope's coordinates x, second it applies the standard X-ray transform on the resulting function. Compactly:

$$X(\phi, x) = X_{std} \circ (x^* \phi), \tag{8.1}$$

where $x^*\phi = x \circ \phi$. We shall describe collections of projection images using the formalism of vector bundles and sections. To this end, consider the vector bundle of images over orientations, $\mathcal{F} \times \mathcal{I}$. In words, the vector bundle attaches to each microscope orientation the vector space of all possible images, we shall denote this vector bundle by $\mathcal{B}$. Matched with $\mathcal{B}$ are its sections, denoted by $\mathcal{H}$. A section $s \in \mathcal{H}$ picks out for each microscope orientation $x \in \mathcal{F}$ a single image from the image space attached to x. Alternatively, sections $s \in \mathcal{H}$ can be viewed as maps $s : \mathcal{F} \to \mathcal{B}$ which fit in the following commutative diagram:

$$\begin{array}{ccc} \mathcal{F} & \xrightarrow{s} & \mathcal{B} \\ & \searrow^{\mathrm{Id}} & \downarrow{\scriptstyle \pi_{\mathcal{F}}} \\ & & \mathcal{F} \end{array}$$

where Id is the identity map and $\pi_{\mathcal{F}}$ is projection on the $\mathcal{F}$ coordinate. We now note that the X-ray transform naturally induces a map, $\mathcal{X} : C_0(B_V) \to \mathcal{H}$,

defined by:

$$(\mathcal{X}\phi)(x) = X(\phi, x),$$

for all $x \in \mathcal{F}$. In words the mapping $\mathcal{X}$ takes a potential function ϕ as input and outputs the section $\mathcal{X}\phi$ which assigns to each orientation x the image given by projecting ϕ from microscope orientation x.

For the remainder of the section we fix a potential function ϕ. The section $\mathcal{X}\phi$ is a continuous object and contains every possible projection of ϕ. However, in practice Cryo-EM experiments collect only a finite number of projection images denoted by $I_1, I_2, ..., I_n$ from unknown orientations. Let the image index set, $\{1, 2, ..., n\}$, be denoted by $[n]$. Then there exists an unknown *orientation map*:

$$\iota : [n] \to \mathcal{F},$$

which for each index $i \in [n]$ returns the microscope orientation used to produce image I_i. We denote the vector bundle of images over $[n]$ by $\widehat{\mathcal{B}}$ and the sections of $\widehat{\mathcal{B}}$ by $\widehat{\mathcal{H}}$. The orientation map induces a restriction map, $\iota^* : \mathcal{H} \to \widehat{\mathcal{H}}$, defined by:

$$(\iota^* s)(i) = s(\iota(i)),$$

for all $i \in [n]$. Putting the definitions together shows the collection of clean images is exactly the section $\iota^*\mathcal{X}(\phi)$. The observed images are a perturbation of this section by noise, $\iota^*\mathcal{X}(\phi) + \epsilon$, models for the noise will be discussed later.

In summary, the relationship between potential functions and observed image collections can be understood in three steps.

$$C_0(B_V) \xrightarrow{\mathcal{X}} \mathcal{H} \xrightarrow{\iota^*} \widehat{\mathcal{H}} \xrightarrow{\epsilon} \widehat{\mathcal{H}}$$

First, the X-ray map, $\mathcal{X}$, converts a potential function ϕ into the continuous image section $\mathcal{X}\phi$ consisting of all projections of ϕ. Next, the restriction map, ι^*, samples the continuous image section $\mathcal{X}\phi$ at the orientations observed by the experiment, producing a discrete image section, $\iota^*\mathcal{X}(\phi)$. Finally, noise is added to give the observed image collection, $\iota^*\mathcal{X}(\phi) + \epsilon$. The geometric idea behind this thesis is the observation that denoising the observed image collection is equivalent to denoising the discrete image section $\iota^*\mathcal{X}(\phi) + \epsilon$.

8.2 Low Dimensional Structure

In this section we identify a finite dimensional subspace of $C_0(B_V)$ which "well contains" potential functions. Our notion of "well contained" depends on the Cryo-EM experiment. Namely, suppose our goal is an r angstrom reconstruction of a potential function ϕ. Then by Nyquist sampling theory we must recover all of the Fourier coefficients of ϕ up to the band limit $b = 1/2r$. Thus, we say a subspace $V_b \subset C_0(B_V)$ well contains potential functions up to resolution $1/2b$ if:

$$P_{V_b}\varphi \approx \varphi_b,$$

for all $\varphi \in C_0(B_V)$, where P_{V_b} is orthogonal projection onto V_b and φ_b is the result of applying a b cutoff filter to φ (i.e. removing all frequencies of φ greater then the band limit b).

For a band limit $b > 0$ we proceed to construct such a subspace using the Fourier-Bessel functions. Our construction closely follows the work of Z. Zhao and A. Singer in [36], which considered a similar problem for the unit disk rather then the unit ball.

The Fourier-Bessel functions are the eigenfunctions of the Laplacian on $C_0(B_V)$ and so give a natural basis. They are defined angularly by the spherical harmonics and radially by the spherical Bessel functions. The spherical harmonics are defined on the unit sphere in V and we denote them using the standard notation, Y_{lm} for all $l \geq 0$ and $|m| \leq l$. The spherical Bessel functions are defined on the unit interval and we denote them using the standard notation, j_l for all $l \geq 0$. Let $x_{l1}, x_{l2}, x_{l3}, ...$ be the zeros of j_l in ascending order. Then the (l, m, k) Fourier-Bessel function, denoted by ψ_{lmk}, is a map $\psi_{lmk} : B_V \to \mathbb{R}$ defined by:

$$\psi_{lmk}(v) = \sqrt{\frac{2}{j_{l+1}^2(x_{lk})}} Y_{lm}(v/||v||) j_l(x_{lk}\,||v||),$$

for all $v \in B_V$. The Fourier-Bessel functions give an orthonormal basis of $C_0(B_V)$. We note that as the index k increases the radial frequency of the Fourier-Bessel functions increases. Likewise, as the index l increases the angular frequency of the Fourier-Bessel functions increases. The relationship between the Fourier-Bessel functions (spherical waves) and the plane waves is given by the following equation:

$$e^{iv\cdot\xi} = \sum_{lmk} S_{lmk}(\xi)\psi_{lmk}(v), \tag{8.2}$$

for all $\xi \in V$ and $v \in B_V$ [33]. Where the expansion coefficients are given by:

$$S_{lmk}(\xi) = 4\pi i^l j_l'(x_{lk}) \frac{j_l(||\xi||)}{||\xi||^2 - x_{lk}^2} Y_{lm}^*(\xi/||\xi||).$$

Equation (8.2) can be used to compute the Fourier transform of the Fourier Bessel functions, giving:

$$(\mathcal{F}\psi_{lmk})(\xi) \propto \frac{j_l(||\xi||)}{||\xi||^2 - x_{lk}^2} Y_{lm}^*(\xi/||\xi||). \tag{8.3}$$

Equation (8.3) shows the Fourier transform of ψ_{lmk} is concentrated about the sphere of radius x_{lk}. Therefore, band limiting functions in $C_0(B_V)$ by b is roughly equivalent to zeroing all Fourier-Bessel coefficients having $x_{lk} \geq b$. To get an estimate for the x_{lk} we use an asymptotic formula for the spherical Bessel functions:

$$j_l(r) \approx \frac{1}{r}\cos\left(r - \frac{\pi(2l+1)}{4} - \frac{\pi}{4}\right). \tag{8.4}$$

Taking the zeros of the approximation gives, $x_{lk} \approx \frac{\pi}{2}(l+2k)$. Using these approximate zeros the requirement $x_{lk} \geq b$ becomes $l + 2k \geq \frac{2b}{\pi}$. Therefore, we take for our finite dimensional subspace:

$$V_b = \bigoplus_{l=0}^{\lfloor \frac{2b}{\pi} - 2 \rfloor} \bigoplus_{m=-l}^{l} \bigoplus_{k=1}^{\lfloor \frac{b}{\pi} - \frac{l}{2} \rfloor} \psi_{lmk}.$$

8.3 Denoising by global projection

The previous chapter identified a subspace $V_b \subset C_0(B_V)$ which well contains potential functions up to the band limit b. Applying the map $\iota^* \circ \mathcal{X}$ to V_b gives the subspace $\iota^*\mathcal{X}(V_b) \subset \widehat{\mathcal{H}}$ which well contains sections of projection images up to the band limit b. Global projection denoises sections of observed images by projecting them onto $\iota^*\mathcal{X}(V_b)$, where projection is defined using the natural inner product on $\widehat{\mathcal{H}}$:

$$\langle s, \widetilde{s} \rangle = \frac{1}{n} \sum_{i=1}^{n} \langle s(i), \widetilde{s}(i) \rangle_{L^2(\mathcal{I})},$$

for all $s, \widetilde{s} \in \widehat{\mathcal{H}}$, where $\langle \cdot, \cdot \rangle_{L^2(\mathcal{I})}$ is the standard L^2 inner product on $\mathcal{I}$.

The SNR of the projected section grows linearly with the number of images. Indeed, suppose we observe a finite collection of images denoted by

$I_1, I_2, ..., I_n$. Each image contributes both signal and noise:

$$I_i = I_i^{clean} + I_i^{noise}.$$

Let σ^2_{signal} be the variance of the clean images and σ^2_{noise} be the variance of the noise images. The collection of clean images gives a section in $\widehat{\mathcal{H}}$ whose energy (norm squared) is approximately σ^2_{signal}. The collection of noise images gives a section in $\widehat{\mathcal{H}}$ whose energy is approximately σ^2_{noise}. Thus, the SNR of the section of observed images (in this thesis SNR is defined to be the energy of the signal divided by the energy of the noise) is approximately $\sigma^2_{signal}/\sigma^2_{noise}$, which is independent of n. However, the energy of the signal is concentrated in the subspace $\iota^*\mathcal{X}(V_b)$ while the energy of the noise is spread equally about $\widehat{\mathcal{H}}$. Therefore, after projecting the section of observed images onto $\iota^*\mathcal{X}(V_b)$ the energy of the signal is still approximately σ^2_{signal}, while the energy of the noise is reduced to approximately $(\dim(\mathrm{V_b})/\mathrm{n})\,\sigma^2_{noise}$. Thus, the SNR of the projection is approximately $(n/\dim(V_b)) \cdot \left(\sigma^2_{signal}/\sigma^2_{noise}\right)$, which grows linearly with the number of input images.

The linear growth of the SNR is formalized in the following theorem. The theorem assumes the orientations of the observed images are chosen from some probability distribution μ_o on $\mathcal{F}$ (not necessarily uniform). The noise images, I^{noise}, are assumed to be i.i.d generated by a probability distribution μ_{noise} on $\mathcal{I}$ which satisfies two conditions. First, we require the noise to be uncorrelated with sections in $\mathcal{X}V_b$, that is:

$$\mathbb{E}\left\langle s(\alpha), I^{noise}\right\rangle = 0,$$

for all $s \in V_b$, where $\alpha \in \mathcal{F}$ is chosen randomly by μ_o and I^{noise} is chosen randomly by μ_{noise}. Second, we require the energy of the noise to be bounded.

That is, there must be a constant $\sigma > 0$ such that $||I^{noise}||$ is always bounded by σ. We note that white noise satisfies these two conditions.

Theorem 7. ***(Linear growth of SNR)*** *Suppose n noisy images of a potential function ϕ are observed, and that the section of observed images is projected onto $\iota^* \mathcal{X} V_b$. Let $\mathcal{E}_\lambda$ be the event that the SNR of the projected images grows linearly:*

$$\mathcal{E}_\lambda = \left(\text{SNR} \geq \frac{n\lambda E_{\phi,b}}{8\dim(V_b)\sigma^2} \right),$$

where $\lambda > 0$ is a constant which controls the rate of the linear growth, and $E_{\phi,b}$ is the expected norm of the signal, $\mathbb{E}\,||P_{\mathcal{X}V_b}\mathcal{X}\phi||$. Then the event $\mathcal{E}_\lambda$ is exponentially likely, that is:

$$\begin{aligned} 1 - \Pr(\mathcal{E}_\lambda) \leq 2\exp\left(\frac{-nE_{\phi,b}^2}{2C_3^2} \right) + \dim(V_b)(.6680)^{\frac{n}{M_b}} \\ + 4\dim(V_b)\exp\left(-\frac{\sigma^2}{\lambda C_1^2 (C_2 + \sigma)^2} \right), \end{aligned}$$

where M_b is a constant depending on b, and the C_is are constants depending on ϕ and b.

For the proof see Appendix C.

Chapter 9

Computing $\iota^*\mathcal{X}(V_b)$

The previous chapter showed that projecting sections of observed images onto $\iota^*\mathcal{X}(V_b)$ gives an effective denoising algorithm. The problem now becomes, how does one compute $\iota^*\mathcal{X}(V_b)$? Recall, that the restriction map ι^* is unknown, therefore the vector space $\iota^*\mathcal{X}(V_b)$ cannot be computed in a straightforward manner. Instead we use the following indirect approach.

1. First, we define V_b in terms of the $SO(V)$ decomposition of $C_0(B_V)$.
2. Next, we show that $\mathcal{X}$ is an $SO(V)$ intertwinner, and identify $\mathcal{X}V_b$ within the $SO(V)$ decomposition of $\mathcal{H}$.
3. Finally, we present a representation theoretic method which calculates an approximation to the ι^* restriction of the $SO(V)$ decomposition of $\mathcal{H}$ to $\widehat{\mathcal{H}}$, within which it is possible to identify a vector space which well contains $\iota^*\mathcal{X}(V_b)$. Observed image collections can then be denoised by projecting onto this slightly larger vector space.

9.1 $SO(V)$ decomposition of $C_0(B_V)$ and $\mathcal{H}$

We first characterize the $SO(V)$ decomposition of $C_0(B_V)$, where $SO(V)$ acts by the left regular action, that is, $g \cdot f = f \circ g^{-1}$ for all $f \in C_0(B_V)$ and

$g \in SO(V)$. A priori, $C_0(B_V)$ has an isotypic $SO(V)$ decomposition:

$$C_0(B_V) = \bigoplus_{l=0}^{\infty} H_l,$$

where each H_l is the sum of multiple copies of the l'th irreducible representation of $SO(V)$. The Laplacian commutes with $SO(V)$ and so it's spectrum refines the decomposition, that is:

$$H_l = \bigoplus_{m=1}^{\infty} H_l^m,$$

for all $l \in \mathbb{N}$, where each H_l^m is an eigenspace of the Laplacian. Therefore, we have:

$$C_0(B_V) = \bigoplus_{l=0}^{\infty} \bigoplus_{m=1}^{\infty} H_l^m \tag{9.1}$$

Theorem 8. *($SO(V)$* ***decomposition of*** *$C_0(B_V)$) Decomposition (9.1) is an irreducible $SO(V)$ decomposition, furthermore V_b is characterized by:*

$$V_b = \bigoplus_{l=0}^{\lfloor \frac{2b}{\pi} - 2 \rfloor} \bigoplus_{m=1}^{\lfloor \frac{b}{\pi} - \frac{l}{2} \rfloor} H_l^m.$$

For a proof see [5]. Next, we characterize the $SO(V)$ decomposition of $\mathcal{H}$, where $SO(V)$ acts by the left regular action, that is, $(g \cdot s)(x) = s(g^{-1}x)$ for all $s \in \mathcal{H}$ and $x \in \mathcal{F}$. A priori $\mathcal{H}$ has an isotypic $SO(V)$ decomposition:

$$\mathcal{H} = \bigoplus_{l=0}^{\infty} E_l.$$

The decomposition is refined using actions and operators which commute with the $SO(V)$ action. To this end, we define two actions, π_{base} and π_{fiber} of $SO(\mathbb{R}^2)$ on $\mathcal{H}$ by:

$$\begin{aligned} (\pi_{base}(h)s)(x) &= s(xh) \\ (\pi_{fiber}(h)s)(x) &= s(x) \circ h, \end{aligned} \tag{9.2}$$

for all $h \in SO(\mathbb{R}^2)$ and $s \in \mathcal{H}$. Geometrically, $\pi_{base}(h)$ permutes the base space $\mathcal{F}$ by multiplication on the right by h, and $\pi_{fiber}(h)$ pointwise rotates image sections by h. Note that the π_{base}, π_{fiber}, and $SO(V)$ actions commute with each other. Therefore, the $SO(V)$ decomposition can be refined, that is:

$$E_l = \bigoplus_{k_b=-\infty}^{\infty} \bigoplus_{k_f=-\infty}^{\infty} E_l^{k_b,k_f},$$

for all $l \in \mathbb{N}$. Where each $E_l^{k_b,k_f}$ component is an irreducible $SO(\mathbb{R}^2)$ representation, characterized by:

$$\begin{aligned} (\pi_{base}(R_\theta)s)(x) &= e^{ik_b\theta}s(x) \\ (\pi_{fiber}(R_\theta)s)(x) &= e^{ik_f\theta}s(x), \end{aligned} \tag{9.3}$$

for all $\theta \in [0, 2\pi]$, where R_θ is counterclockwise rotation by θ. To further refine the decomposition consider the unbounded operator $\Delta_{fiber} : \mathcal{H} \to \mathcal{H}$ which pointwise applies the two dimensional Laplacian to image sections:

$$(\Delta_{fiber}s)(x) = \Delta_{\mathbb{R}^2}s(x),$$

for all $x \in \mathcal{F}$. Because Δ_{fiber} acts pointwise, it commutes with the $SO(V)$ and π_{base} actions. Furthermore, since two dimensional rotations commute with the two dimensional Laplacian we have that Δ_{fiber} commutes with π_{fiber}. Therefore, the spectrum of Δ_{fiber} further refines the decomposition, that is:

$$E_l^{k_b,k_f} = \bigoplus_{q=1}^{\infty} E_l^{k_b,k_f,q},$$

for all $l \in \mathbb{N}$ and $k_b, k_f \in \mathbb{Z}$. Where each component $E_l^{k_b,k_f,q}$ is an eigenspace of Δ_{fiber} with corresponding eigenvalue $\lambda_l^{k_b,k_f,q}$. Without loss of generality we assume the eigenvalues are increasing in q. Combining the refinements gives:

$$\mathcal{H} = \bigoplus_{l=0}^{\infty} \bigoplus_{k_b=-\infty}^{\infty} \bigoplus_{k_f=-\infty}^{\infty} \bigoplus_{q=1}^{\infty} E_l^{k_b,k_f,q}. \tag{9.4}$$

The following theorem, proved in Appendix C, characterizes the above decomposition in terms of eigenfunctions of $\Delta_{\mathbb{R}^2}$ on $\mathcal{I}$. The eigenfunctions are denoted by $\psi^{k,q}$ for $k \in \mathbb{Z}$ and $q \in \mathbb{N}^+$. The corresponding eigenvalues are denoted $\mu^{k,q}$, and for fixed k increase in q, that is, $\mu^{k,1} < \mu^{k,2} < ...$ In polar coordinates the eigenfunctions are given by:

$$\psi^{k,q}(\theta, r) = \frac{1}{\sqrt{\pi}|J_{k+1}(x_{kq})|} J_k(z_{kq} r) e^{ik\theta},$$

where J_k is the Bessel function of order k and z_{kq} is the qth zero of J_k.

Theorem 9. *$SO(V)$ **decomposition of** $\mathcal{H}$*

1. *$E_l^{k_b,k_f,q} = 0$ for all $|k_b| > l$.*
2. *$\lambda_l^{k_b,k_f,q} = \mu^{k_f,q}$.*
3. *Each $s \in E_l^{k_b,k_f,q}$ is fiberwise a scalar multiple of $\psi^{k_f,q}$.*
4. *$E_l^{k_b,k_f,q}$ is irreducible for each l, k_b, k_f, q.*

9.2 Characterizing $\mathcal{X}$

Section 9.1 characterized the $SO(V)$ decomposition of both the space of potential function and the space of image sections. This section studies how the X-ray mapping, $\mathcal{X}$, relates the two decompositions. In particular we want to identify $\mathcal{X}V_b$ within the $SO(V)$ decomposition of $\mathcal{H}$. To this end, the following theorem lists the symmetries of $\mathcal{X}$.

Theorem 10. ***Symmetries of** $\mathcal{X}$*

1. *$\mathcal{X}$ intertwines the $SO(V)$ actions.*

2. On the image of $\mathcal{X}$ the actions π_{base} and π_{fiber} are equivalent. That is:

$$\pi_{base}(h)(\mathcal{X}\phi) = \pi_{fiber}(h)(\mathcal{X}\phi),$$

for all $\phi \in C_0(B_V)$ and $h \in SO(\mathbb{R}^2)$.

For the proof see Appendix C. Theorem 10.1 gives that $\mathcal{X}$ is an $SO(V)$ intertwinner, therefore, because decomposition (9.1) and (9.4) are irreducible we can invoke Shur's lemma to give that:

$$\pi_{l'}^{k_b,k_f,q} \circ \mathcal{X} : H_l^m \to E_{l'}^{k_b,k_f,q}, \tag{9.5}$$

is zero for all $l' \neq l$, where $\pi_{l'}^{k_b,k_f,q}$ is orthogonal projection onto $E_{l'}^{k_b,k_f,q}$. Next we invoke the equivariance property given by Theorem 10.2 to see that:

$$\pi_{l'}^{k_b,k_f,q} \circ \mathcal{X} : H_l^m \to E_{l'}^{k_b,k_f,q} \tag{9.6}$$

is zero for all $k_b \neq k_f$. By applying equations (9.5) and (9.6) we can describe $\mathcal{X}V_b$:

$$\mathcal{X}(V_b) = \bigoplus_{l=0}^{\lfloor \frac{2b}{\pi} - 2 \rfloor} \bigoplus_{q=1}^{\lfloor \frac{b}{\pi} - \frac{l}{2} \rfloor} \mathcal{X}(H_l^m) \subseteq \bigoplus_{l=0}^{\lfloor \frac{2b}{\pi} - 2 \rfloor} \bigoplus_{|k| \leq l} \bigoplus_{q=1}^{\infty} E_l^{k,k,q}.$$

Let W_b denote the vector space on the right. It is preferable to $\mathcal{X}V_b$ in that it is defined solely in terms of the $SO(V)$ decomposition of $\mathcal{H}$. We will later show that this allows an ι^* restriction of it to be computed using spectral methods. Also, because $\mathcal{X}V_b$ is contained in W_b we have that W_b also well contains Cryo-EM image sections. Therefore, observed Cryo-EM image sections can be denoised by projecting onto $\iota^* W_b$ instead of $\iota^* \mathcal{X}(V_b)$. A problem arise though in that W_b is infinite dimensional. To have an effective denoising algorithm with SNR scaling linearly with the number of input images we must make a

truncation decision on W_b. That is we need to choose $Q_{l,k} \in \mathbb{N}$ such that $\mathcal{X}V_b$ is well contained in the finite dimensional space:

$$\widetilde{W_b} = \bigoplus_{l=0}^{\lfloor \frac{2b}{\pi} - 2 \rfloor} \bigoplus_{|k| \leq l} \bigoplus_{q=1}^{Q_{l,k}} E_l^{k,k,q}.$$

This will be done using the Fourier slice theorem which states that the $2D$ Fourier transform of projection images are equivalent to restrictions of the $3D$ Fourier transform of potential functions.

Theorem 11. ***(Fourier Slice Theorem)*** *Let $\mathcal{FT}_V$, $\mathcal{FT}_{\mathbb{R}^2}$ denote the Fourier transform on $C_0(B_V)$ and $C_0(D_{\mathbb{R}^2})$ respectively, then:*

$$(\mathcal{FT}_V \phi) \circ x \circ \iota_{1,2} = \mathcal{FT}_{\mathbb{R}^2}(X(\phi, x)),$$

for all $\phi \in C_0(B_V)$ and $x \in \mathcal{F}$, where $\iota_{1,2} : \mathbb{R}^2 \to \mathbb{R}^3$ the standard inclusion map.

For a proof see [17]. Theorem 11 implies that if a potential function ϕ is band limited by b, then all of its projection images will also be band limited by b. Because our goal is to find a finite dimensional subspace $\widetilde{W_b}$ which well contains $\mathcal{X}V_b$ this means we can throw out any components $E_l^{k,k,q}$ whose pointwise 2D Fourier transform's are concentrated outside the band limit b. To this end, recall that theorem 9.3 states that sections in $E_l^{k,k,q}$ are pointwise scalar multiples of the 2D Fourier-Bessel function $\psi^{k,q}$. A calculation similar to the one performed in section 8.2 computes the Fourier transform of the 2D Fourier-Bessel functions, giving:

$$(\mathcal{FT}_{\mathbb{R}^2} \psi^{k,q})(r, \theta) \propto \frac{J_k(2\pi r)}{(2\pi r)^2 - z_{kq}^2} e^{ik\theta}.$$

For details see [36]. Therefore, the Fourier transform of $\psi^{k,q}$ is concentrated about the circle of radius $\frac{z_{kq}}{2\pi}$. We only want Fourier-Bessel functions whose Fourier transform is concentrated within the band limit b so the cutoff criteria is $z_{kq} \leq 2\pi b$. An asymptotic formula for the Bessel zeros gives:

$$z_{kq} \approx \frac{\pi}{2}(k + 2q - 1).$$

Using these approximate zeros the cutoff criteria becomes: $q \leq \frac{b}{\pi} - \frac{k}{2} + \frac{1}{2}$, for fixed $k \in \mathbb{Z}$. Denote the floor of the cutoff criteria by Q_k. Then we define $\widetilde{W_b}$ by:

$$\widetilde{W_b} = \bigoplus_{l=0}^{\lfloor \frac{2b}{\pi} - 2 \rfloor} \bigoplus_{|k| \leq l} \bigoplus_{q=1}^{Q_k} E_l^{k,k,q}. \tag{9.7}$$

The subspace $\widetilde{W_b}$ has three important properties. First, it is defined entirely in terms of representation theory. Second, it is finite dimensional. Finally, it well contains $\mathcal{X}V_b$. Therefore, we can denoise observed image sections by projecting onto $\iota^*\widetilde{W_b}$ rather then $\iota^*\mathcal{X}(V_b)$.

9.3 Computing $\iota^*\widetilde{W_b}$

In this section we present a procedure which computes an approximation to $\iota^*\widetilde{W_b}$. The previous section characterized $\widetilde{W_b}$ in terms of the π_{fiber}, $\Delta_{\mathbb{R}^2}$, and $SO(V)$ decompositions of $\mathcal{H}$. The first two actions are defined fiberwise on $\mathcal{H}$ and so have natural analogs on $\widehat{\mathcal{H}}$, which induce a decomposition:

$$\widehat{\mathcal{H}} = \bigoplus_{k_f=-\infty}^{\infty} \bigoplus_{q=1}^{\infty} A^{k_f,q}, \tag{9.8}$$

where analogous to theorem 9 it can be shown sections $s \in A^{k_f,q}$ are pointwise scalar multiples of the $\psi^{k_f,q}$ eigenfunction of $\Delta_{\mathbb{R}^2}$. Applying the restriction

map ι^* to $\widetilde{W}_b$ gives:

$$\begin{aligned}\iota^*\widetilde{W}_b &= \bigoplus_{|k|\leq\lfloor\frac{2b}{\pi}-2\rfloor}\bigoplus_{q=1}^{Q_k}\iota^*\left(\bigoplus_{l=|k|}^{\lfloor\frac{2b}{\pi}-2\rfloor}E_l^{k,k,q}\right)\\ &\subset \bigoplus_{|k|\leq\lfloor\frac{2b}{\pi}-2\rfloor}\bigoplus_{q=1}^{Q_k}A^{k,q},\end{aligned}$$

where containment follows from the fact that ι^* commutes with the π_{fiber} and Δ_{fiber} actions. Therefore, computing $\iota^*\widetilde{W}_b$ is equivalent to computing within $A^{k,q}$ the subspace $A_b^{k,q}$, defined by:

$$A_b^{k,q} = \bigoplus_{l=|k|}^{\lfloor\frac{2b}{\pi}-2\rfloor}\iota^*E_l^{k,k,q},$$

for all $|k| \leq \lfloor\frac{2b}{\pi}-2\rfloor$ and $q \leq Q_k$. A problem arises in that the subspace is defined by the $SO(V)$ decomposition of $\mathcal{H}$, because the orientations of the observed projection images are unknown there is no meaningful $SO(V)$ action on $\widehat{\mathcal{H}}$ and so we cannot directly compute $A_b^{k,q}$ through decomposition. Instead, we indirectly perform an $SO(V)$ decomposition of $\widehat{\mathcal{H}}$ in the following way. First, we identify an operator $T : \mathcal{H} \to \mathcal{H}$ which respects the actions used to decompose $\mathcal{H}$ and can be computed from the observed images. Second, we demonstrate how to compute a restriction of the operator to $\widehat{\mathcal{H}}$. Finally, we demonstrate how the spectrum of the restricted operator gives an approximate $SO(V)$ decomposition of $\widehat{\mathcal{H}}$.

9.3.1 Class averaging operator

Consider for each pair of orientations $x, y \in \mathcal{F}$ the rotation that best aligns the projection image from orientation y with the projection image from

orientation x. We call this rotation the *alignment kernel* and denote it by $\widehat{k}(x,y)$, then:

$$\begin{aligned}\widehat{k}(x,y) &= \underset{g\in SO(\mathbb{R}^2)}{\operatorname{argmin}} ||g\cdot(\mathcal{X}\phi)(y)-(\mathcal{X}\phi)(x)|| \\ &= \underset{g\in SO(\mathbb{R}^2)}{\operatorname{argmin}} \left|\left|(\mathcal{X}\phi)(y\circ g^{-1})-(\mathcal{X}\phi)(x)\right|\right| \\ &\approx \underset{g\in SO(\mathbb{R}^2)}{\operatorname{argmin}} \left|\left|y\circ g^{-1}-x\right|\right|_{\mathrm{F}},\end{aligned} \tag{9.9}$$

where F denotes the Frobenius norm. The first equality follows from definitions, the approximation follows from the continuity of $(\mathcal{X}\phi)(\cdot)$. The alignment kernel was introduced by A. Singer, Z. Zhao, Y. Shkolnisky, and R. Hadani in [24] and [12] where it was shown that:

$$\pi_{1,2}\circ x^{-1}\circ R_{x_3,y_3}\circ y\circ \iota_{1,2} = \underset{g\in SO(\mathbb{R}^2)}{\operatorname{argmin}} \left|\left|y\circ g^{-1}-x\right|\right|_{\mathrm{F}}, \tag{9.10}$$

where $\iota_{1,2}:\mathbb{R}^2\to\mathbb{R}^3$ is the standard embedding, $\pi_{1,2}:\mathbb{R}^3\to\mathbb{R}^2$ is the standard projection, and R_{x_3,y_3} is the smallest rotation taking $y(e_3)$ to $x(e_3)$. We call the left hand side of equation (9.10) the *class averaging kernel* and denote it by $k(x,y)$. Equation (9.9) shows that the alignment kernels gives an approximation of the class averaging kernel. The class averaging kernels define an operator, $\mathcal{C}:\mathcal{H}\to\mathcal{H}$, given by:

$$(\mathcal{C}s)(x) = \frac{1}{|N_x|}\int_{y\in N_x} k(x,y)\cdot s(y)\,\mathrm{d}\mu_o(y),$$

where N_x is a collection of orientations whose viewing directions are close to the viewing direction of x, and μ_0 is a probability measure on $\mathcal{F}$. We call the operator the *class averaging operator*. Geometrically, it takes as input sections $s\in\mathcal{H}$ and outputs a new section $\mathcal{C}s$ which at each $x\in\mathcal{F}$ is computed by taking nearby images, aligning them to $s(x)$, and averaging. We note that denoising

by class averaging amounts to constructing a class averaging operator and applying it to the section of observed images. The following theorem lists the group actions which the class averaging operator preserves.

Theorem 12. ***Symmetries of*** $\mathcal{C}$

1. $\mathcal{C}$ *intertwines the* π_{fiber} *action.*

2. *If* μ_0 *is invariant under right multiplication by* $SO(\mathbb{R}^2)$*, then* $\mathcal{C}$ *intertwines the* π_{base} *action.*

3. *If* μ_0 *is invariant under right multiplication by* $SO(\mathbb{R}^2)$*, then* $\mathcal{C}$ *intertwines the* π_{base} *action with the* π_{fiber} *action.*

4. $\mathcal{C}$ *commutes with* Δ_{fiber}*.*

5. *If* μ_0 *is* $SO(V)$ *invariant and the neighborhoods are invariant in the sense that,*

$$gN_x = N_{gx}$$

for all $x \in \mathcal{F}$ *and* $g \in SO(V)$*, then* $\mathcal{C}$ *intertwines the* $SO(V)$ *action.*

The proof follows from the geometric definition of the class averaging kernel, for the details see Appendix C. We note that the $SO(\mathbb{R}^2)$ invariance of μ_0 can be enforced by randomly rotating the observed images, for the remainder of the thesis we assume this has been done. We now use the symmetries of $\mathcal{C}$ to characterize it's spectrum. First, note that the $SO(\mathbb{R}^2)$ and Δ_{fiber} components of $\mathcal{H}$ are invariant under $\mathcal{C}$, that is:

$$\mathcal{C} : \bigoplus_{l=0}^{\infty} E_l^{k_b,k_f,q} \to \bigoplus_{l=0}^{\infty} E_l^{k_b,k_f,q},$$

for all $k_b, k_f \in \mathbb{Z}$ and $q \in \mathbb{N}^+$. This follows from items one, two, and six of theorem 12 combined with Shur's lemma. Next, note that $\mathcal{C}$ is zero on components with different π_{base} and π_{fiber} indices, that is:

$$\mathcal{C} : \bigoplus_{l=0}^{\infty} E_l^{k_b,k_f,q} \to 0,$$

for all $k_b \neq k_f$. This follows from item three of theorem 12. Finally, note that if the hypothesis of item five of theorem 12 is satisfied then Schur's lemma gives that the irreducible $SO(V)$ components are invariant under $\mathcal{C}$, that is:

$$\mathcal{C} : E_l^{k_b,k_f,q} \to E_l^{k_b,k_f,q},$$

with $\mathcal{C}$ equal to a scalar multiple of the identity.

9.3.2 Restricting the class averaging operator

We now show how a restriction of $\mathcal{C}$ can be used to compute an approximation of $\iota^*\widetilde{W_b}$. Suppose we observe a finite collection of projection images from a Cryo-EM experiment denoted $I_1, I_2, ..., I_n$. Using established (though highly non-trivial) Cryo-EM techniques, we can identify pairs of images which are similar up to translation, roatation, and nosie [8], [37]. For such pairs of images we can compute an alignment kernel. Combining the alignment kernels together we can form an empirical class averaging operator, $\widehat{\mathcal{C}} : \widehat{\mathcal{H}} \to \widehat{\mathcal{H}}$, defined by:

$$(\widehat{\mathcal{C}}s)(i) = \frac{1}{|N_i|} \sum_{j \in N_i} \widehat{k}(i,j) \cdot s(j),$$

where N_i denotes the collection of indices with corresponding images close to I_i and $\widehat{k}(i,j)$ is the alignment kernel between images I_j and I_i. We note that $\widehat{\mathcal{C}}$ commutes with the actions of π_{fiber} and Δ_{fiber} on $\widehat{\mathcal{H}}$, so we can restrict it

to components of decomposition (9.8):

$$\widehat{\mathcal{C}} : A^{f,q} \to A^{k,q},$$

for all $k \in \mathbb{Z}$ and $q \in \mathbb{N}^+$. Suppose that the orientations are distributed uniformly and the neighborhoods N_i are the ι^* restriction of $SO(V)$ invariant neighborhoods (in the sense of theorem 12.5). Then restricting $\widehat{\mathcal{C}}$ to $A^{k,q}$ and taking the span of the top $(\lfloor \frac{2b}{\pi} - 2 \rfloor)^2 - (|k| - 1)^2$ eigenvectors produces a vector space which we denote by $\widehat{A}_b^{k,q}$. By Schur's lemma we have that as a vector space $\widehat{A}_b^{k,q}$ is close to $A_b^{k,q}$, with the approximation growing tighter with increasing n.

Unfortunately, the hypothesis of 12.5 does not hold in practice. However, we argue that if we oversample the top eigenspaces of $\widehat{\mathcal{C}}$ restricted to $A^{k,q}$, then $A_b^{k,q}$ will be well contained in their span. To see the intuition of why this is true note that $\widehat{\mathcal{C}}$ is an averaging operator and that sections $s \in A_b^{k,q}$ are smooth. Therefore, the energy term:

$$\left\langle \widehat{\mathcal{C}} s, s \right\rangle,$$

is large. Thus s is well contained in the span of the top eigenspaces of $\widehat{\mathcal{C}}$. The following theorem makes this more precise. The theorem concerns an finite dimensional inner product vector space V and a symmetric operator $T : V \to V$. Let $\lambda_1 > \lambda_2 > ... > \lambda_n$ be the eigenvalues of T with corresponding eigenspace $V_1, V_2, ..., V_n$.

Theorem 13. ***(Davis bound for high energy vector spaces)*** *Suppose W is a subspace of V where for all non-zero $v \in W$:*

$$\langle Tv, v \rangle \geq E \left|\left| v \right|\right|^2, \tag{9.11}$$

for some constant $E > 0$. Let S_m denote the span of the top m eigenspace of T, then:

$$\left|\left|P_{S_m^{\perp}} P_W\right|\right| \leq \sqrt{\frac{\lambda_1 - E}{\lambda_1 - \lambda_{m+1}}}, \tag{9.12}$$

where we use the operator norm, and P_A denotes orthogonal projection onto a subspace A.

For the proof see Appendix C. For noise-free images theorem 13 shows how $A_b^{k,q}$ can be made arbitrarily well contained within the spectrum of an empirical class averaging operator. Namely, the neighborhoods over which $\widehat{\mathcal{C}}$ averages can be made very small thus causing both E and λ_1 to approach one and so the numerator of the bound in equation (9.12) to approach zero. Also, $\widehat{\mathcal{C}}$ is a discretization of a compact operator, so it's spectrum approaches zero. Therefore, by taking enough eigenvectors the denominator of the bound in equation (9.12) approaches one. Combining these approaches together the bound in equation (9.12) can be taken to zero. In practice the the images are corrupted by noise, which limits our ability to identify small neighboorhoods. In this case while oversampling eigenvectors does not take the error bound to zero, it does make the error bound small.

Chapter 10

Shifts

The images collected from Cryo-EM experiments are not centered. That is, the observed images are shifts of projection images. We parameterize shifts by pairs $(s_x, s_y) \in \mathbb{R}^2$ which specify how far the image is shifted in the x and y direction respectively. Specifically, for an image I we define it's (s_x, s_y) shift by:

$$[(s_x, s_y) \cdot I](a, b) = I(a - s_x, b - s_y),$$

for all $(a, b) \in D_{\mathbb{R}^2}$. Recall, in section 8.1 we modeled observed image collections as section in $\widehat{\mathcal{H}}$. To describe shifts of image sections we introduce the vector space of *shift sections*.

Definition 8. *A **shift section** is a mapping $v : [n] \to \mathbb{R}^2$. Shift sections act on discrete image sections by:*

$$(v \cdot s)(i) = v(i) \cdot s(i),$$

for all $s \in \widehat{\mathcal{H}}$ and $i \in [n]$.

In this chapter we look at the effect global projection has when applied to sections of shifted images.

10.1 Global projection of shifted image sections

Recall, that global projection denoises image sections by projection onto a subspace $W \subset \widehat{\mathcal{H}}$ which well contains the signal. We now model image sections from a potential function ϕ by:

$$v \cdot (\iota^* \mathcal{X})(\phi) + \epsilon,$$

where v is an unknown shift section chosen from a probability distribution. Taking the expectation (under the shift section distribution) of projection onto W gives:

$$\begin{aligned} \mathbb{E}[P_W(v \cdot (\iota^* \mathcal{X})(\phi) + \epsilon)] &= \mathbb{E}[P_W v \cdot (\iota^* \mathcal{X})(\phi)] + P_W \epsilon \\ &= P_W \mathbb{E}[v \cdot (\iota^* \mathcal{X})(\phi)] + P_W \epsilon \\ &\approx \mathbb{E}[v \cdot (\iota^* \mathcal{X})(\phi)], \end{aligned} \tag{10.1}$$

where the second equality follows from the linearity of expectation and the approximation follows from the fact that W well contains signal but not noise. The shift sections for observed Cryo-EM image collections are well approximated by a random model which independently chooses the shift of each index from a two dimensional Gaussian centered at the origin:

$$v(i) \sim N((0,0), \sigma^2 I), \tag{10.2}$$

for all $i \in [n]$. Suppose a shift section w is choosen from such a distribution, then the approximation:

$$\mathbb{E}[P_W(v \cdot (\iota^* \mathcal{X})(\phi) + \epsilon)] \approx P_W(w \cdot (\iota^* \mathcal{X})(\phi) + \epsilon), \tag{10.3}$$

follows from the law of large numbers. Combining equations (10.1) and (10.3) gives:

$$P_W(w \cdot (\iota^* \mathcal{X})(\phi) + \epsilon) \approx \mathbb{E}[v \cdot (\iota^* \mathcal{X})(\phi)], \tag{10.4}$$

where expectation is taken with respect to distribution (10.2). We note that the right hand side of equation (10.4) blurs image sections by:

$$\mathbb{E}[v \cdot (\iota^* \mathcal{X})(\phi)](i) = \mathbb{E}[(s_x, s_y) \cdot (\iota^* \mathcal{X})(\phi)(i)],$$

for all $i \in [n]$, where expectation on the right is taken with respect to a two dimensional Gaussian. Figure 10.1 shows the blurring effect applied to a simulated projection image. We note that the blurring can be undone by aligning the observed projection images onto the blurred (but centered) images, and applying global projection to the resulting images.

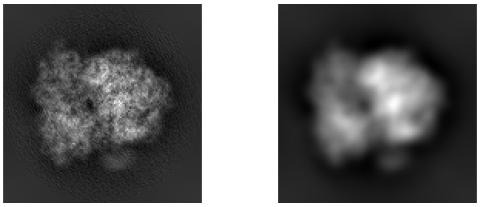


Figure 10.1: Standard deviation six angstrom Gaussian blurring of a clean projection image.

In summary: projecting a section of non-centered projection images onto W returns a section of blurred projection images. *The key point is that the output section is centered.* Therefore, projection both denoises and centers image collections.

Chapter 11

Implementing global projection

In this chapter we make a coordinate choice for $\widehat{\mathcal{H}}$ and express the class averaging operator and global projection in terms of this coordinate choice. Using these expressions we outline an implementation of the global projection algorithm.

11.1 Fourier-Bessel coordinates of $\widehat{\mathcal{H}}$

Recall the decomposition:

$$\widehat{\mathcal{H}} = \bigoplus_{k=-\infty}^{\infty} \bigoplus_{q=1}^{\infty} A^{k,q},$$

where sections $s \in A^{k,q}$ are pointwise scalar multiples of the $\psi^{k,q}$ eigenfunction of $\Delta_{\mathbb{R}^2}$. Consider the maps:

$$a^{k,q} : A^{k,q} \to \mathbb{C}[n],$$
$$b^{k,q} : \mathbb{C}[n] \to A^{k,q},$$

defined by:

$$(a^{k,q}s)(i) = \left\langle s(i), \psi^{k,q} \right\rangle,$$
$$(b^{k,q}f)(i) = f(i)\psi^{k,q},$$

for all $i \in [n]$. Then $a^{k,q}$ and $b^{k,q}$ are inverses of each other and isometries. In words the map $a^{k,q}$ takes as input a section of images and outputs a vector

consisting of the (k, q) Fourier-Bessel coefficients of the images. Combining the $a^{k,q}$ mappings together defines an isometry a with:

$$a : \widehat{\mathcal{H}} \to \bigoplus_{k=-\infty}^{\infty} \bigoplus_{q=1}^{\infty} \mathbb{C}[n].$$

We call the pair $\left(a, \bigoplus_{k=-\infty}^{\infty} \bigoplus_{q=1}^{\infty} \mathbb{C}[n]\right)$ the Fourier-Bessel coordinates of $\widehat{\mathcal{H}}$.

11.1.1 Class averaging operator in coordinates

Recall the components $A^{k,q}$ are invariant under the class averaging operator. The restriction of the class averaging operator to $A^{k,q}$ can be expressed in the Fourier-Bessel coordinate system:

$$a^{k,q} \circ \widehat{\mathcal{C}} \circ b^{k,q} : \mathbb{C}[n] \to \mathbb{C}[n].$$

We denote this operator by $C^{k,q}$, then:

$$\begin{aligned}
(C^{k,q} f)(i) &= a^{k,q} \left(\frac{1}{|N(i)|} \sum_{j \in N(i)} k(i,j) \cdot (b^{k,q} f)(j) \right) \\
&= a^{k,q} \left(\frac{1}{|N(i)|} \sum_{j \in N(i)} k(i,j) \cdot \psi^{k,q} f(j) \right) \\
&= a^{k,q} \left(\psi^{k,q} \frac{1}{|N(i)|} \sum_{j \in N(i)} \exp(\sqrt{-1} k \theta_{i,j}) f(j) \right) \\
&= \frac{1}{|N(i)|} \sum_{j \in N(i)} \exp(\sqrt{-1} k \theta_{i,j}) f(j),
\end{aligned}$$

for all $f \in \mathbb{C}[n]$, where $\theta_{i,j}$ is angle of the rotation which aligns image I_j to I_i. We note that the operator $C^{k,q}$ does not depend on q, therefore we shall henceforth denote the operator C^k.

The eigenvectors of $\widehat{\mathcal{C}}$ and C^k are related by the $a^{k,q}$ maps. Namely, let $V_m^{k,q}$ denote the span of the top m eigenvectors of $\widehat{\mathcal{C}}$ restricted to $A^{k,q}$. Let W_m^k denote the span of the top m eigenvectors of C^k, then:

$$a^{k,q}\left(V_m^{k,q}\right) = W_m^k.$$

We also note that because $C_{-k} = \overline{C_k}$ we have $V_m^{-k} = \overline{V_m^k}$.

11.1.2 Global projection in coordinates

Fix a band limit b. Let $V_m^{k,q}$ denote the span of the top m eigenvectors of $\widehat{\mathcal{C}}$ restricted to $A^{k,q}$. Define a vector space $V_m \subset \widehat{\mathcal{H}}$ by:

$$V_m = \bigoplus_{|k|\leq k_{max}} \bigoplus_{q=1}^{Q_k} V_m^{k,q},$$

where $k_{max} = \lfloor \frac{2b}{\pi} - 2 \rfloor$ and $Q_k = \lfloor \frac{b}{\pi} - \frac{k}{2} + 1 \rfloor$. Then for large m the subspace V_m well contains the section of clean images. Let P_{V_m} denote projection onto V_m, then:

$$\begin{aligned} P_{V_m} &= \sum_{|k|\leq k_{max}} \sum_{q=1}^{Q_k} P_{V_m^{k,q}} \\ &= \sum_{|k|\leq k_{max}} \sum_{q=1}^{Q_k} b^{k,q} \circ P_{W_m^k} \circ a^{k,q} \end{aligned}$$

In words, projecting a section of images onto V_m is equivalent to projecting the vector consisting of the (k,q) Fourier-Bessel coefficients of the images onto the subspace W_m^k (for each pair (k,q) within the bounds established by the band limit b), and then reconstructing the section of images using these denoised Fourier-Bessel coefficients.

11.2 Implementation outline

We can now outline an implementation of global projection. The following table lists the input of the algorithm.

Notation	Description
$I_1, I_2, ..., I_n$	Noisy input images.
b	Band limit of the Fourier coefficients to denoise.
n_{nbh}	Number of nearby neighbors used to construct the class averaging operator.
m	Number of eigenvectors to take from restrictions of the class averaging operator.

Table 11.1: Global projection inputs.

Algorithm outline:

1. For each image I_i identify a collection of n_{nbh} similar images (up to translation, rotation, and noise). Denote the set of these image's indicies by $N(i)$.

2. Identify the angle $\theta_{i,j}$ which optimally aligns image I_j to image I_i for all pairs (i, j) with $j \in N(i)$.

3. For each k with $0 \leq k \leq \lfloor \frac{2b}{\pi} - 2 \rfloor$ construct a sparse matrix $C^k \in \mathbb{C}^{n \times n}$ by:
$$C^k(i,j) = \begin{cases} \exp(\sqrt{-1}k\theta_{i,j}) & \text{if} \quad j \in N(i) \\ 0 & \text{if} \quad j \notin N(i) \end{cases}$$
Take the top m left singular vectors of C^k. Denote the matrix consisting of these vectors by $W_m^k \in \mathbb{C}^{n \times m}$.

4. Construct vectors $c^{k,q} \in \mathbb{C}^{n\times 1}$ by:

$$c^{k,q}(i) = \left\langle I_i, \psi^{k,q} \right\rangle,$$

for all pairs (k,q) with $|k| \leq k_{max} \equiv \lfloor \frac{2b}{\pi} - 2 \rfloor$ and $1 \leq q \leq q^k_{max} \equiv \lfloor \frac{b}{\pi} - \frac{k}{2} + 1 \rfloor$.

5. Construct vectors $\widetilde{c}^{k,q} \in \mathbb{C}^{n\times 1}$ by:

$$\widetilde{c}^{k,q} = \begin{cases} W^{k,q}(W^{k,q})^* c^{k,q} & \text{if} \quad k \geq 0 \\ \overline{W^{k,q}}(W^{k,q})^\intercal c^{k,q} & \text{if} \quad k < 0 \end{cases}$$

for all pairs (k,q) with $|k| \leq k_{max}$ and $1 \leq q \leq q^k_{max}$.

6. Construct denoised images $\widetilde{I}_1, \widetilde{I}_2, ..., \widetilde{I}_n$ by:

$$\widetilde{I}_i = \sum_{|k|\leq k_{max}} \sum_{q=1}^{q^k_{max}} \widetilde{c}^{k,q}(i)\psi^{k,q}.$$

Notes on the algorithm:

- The most computationally intensive part of the algorithm is the first step. The navie approach to identifying nearest neighbors (comparing every pair of images) is complexity $O(n^2)$, which is prohibitively large for large image collections. A fast approach was developed by Z. Zhao and A. Singer in [37]. Briefly: first rotationally invariant features are extracted from the images, then using these extracted features a randomized algorithm quickly identifies nearest neighbors.

- The class averaging operators in step 3 can be given different normalizations. For example using the standard graph Laplacian normalization gives:

$$C^k_{normalized}(i,j) = \frac{1}{\sqrt{|N(i)||N(j)|}} C^k(i,j).$$

Chapter 12

Numerical results

In this chapter we present numerical results from applying the global projection algorithm to real and simulated Cryo-EM data sets. For comparison, the data sets were also denoised using class averaging as implemented in the ASPIRE package by Z. Zhao, A. Singer, and Y. Shkolnisky [37].

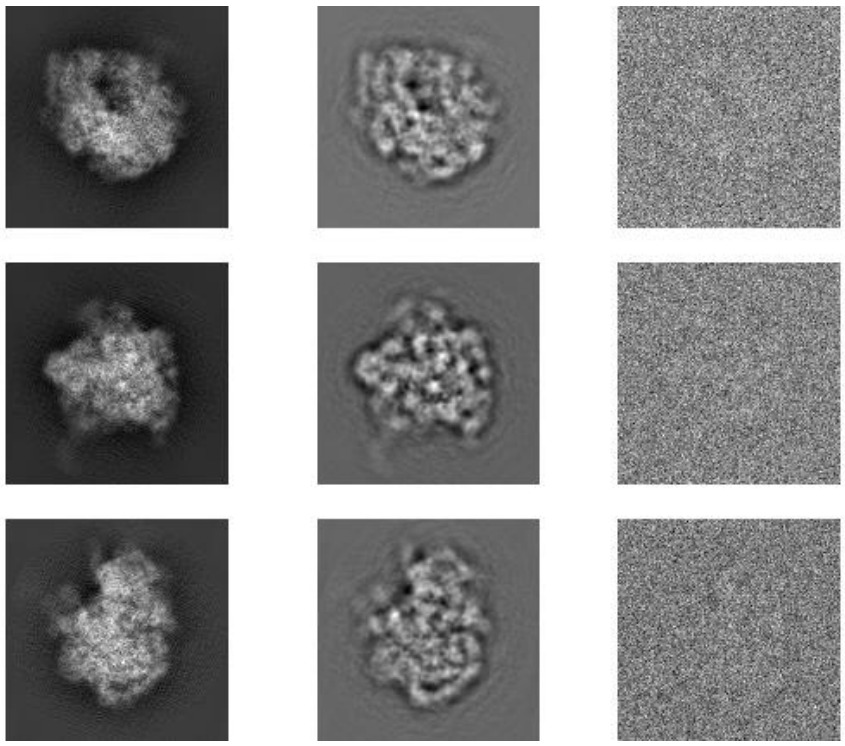
12.1 Simulated data sets

The global projection algorithm was applied to two sets of simulated projection images of the 50s ribosome subunit. The first image set consisted of high contrast images, the second image set consisted of low contrast images. For the high contrast images the two methods exhibited similar performance. For the low contrast images global projection outperformed class averaging.

12.1.1 High contrast images

The ASPIRE package was used to generate 2.5×10^4 simulated images of a 50s ribosome subunit, each image had 129×129 pixels with a pixel length of 2.82 Å. The image collection consisted of 20 defocus groups with the defocus ranging from 1.5 μm to 3.0 μm (defocus is a microscope parameter which controls the contrast of the images, higher defocus values yield images with better contrast. A defocus group is a collection of images collected using the

same microscope defocus. For a brief overview of defocus see appendix C). White noise was then added to the images to give SNR = 1/100. Next, the images were phase flipped (this partially corrects for the defocus, see appendix C for details). Finally, the images were given a random shift between -5 to $+5$ pixels in both the x and y directions.

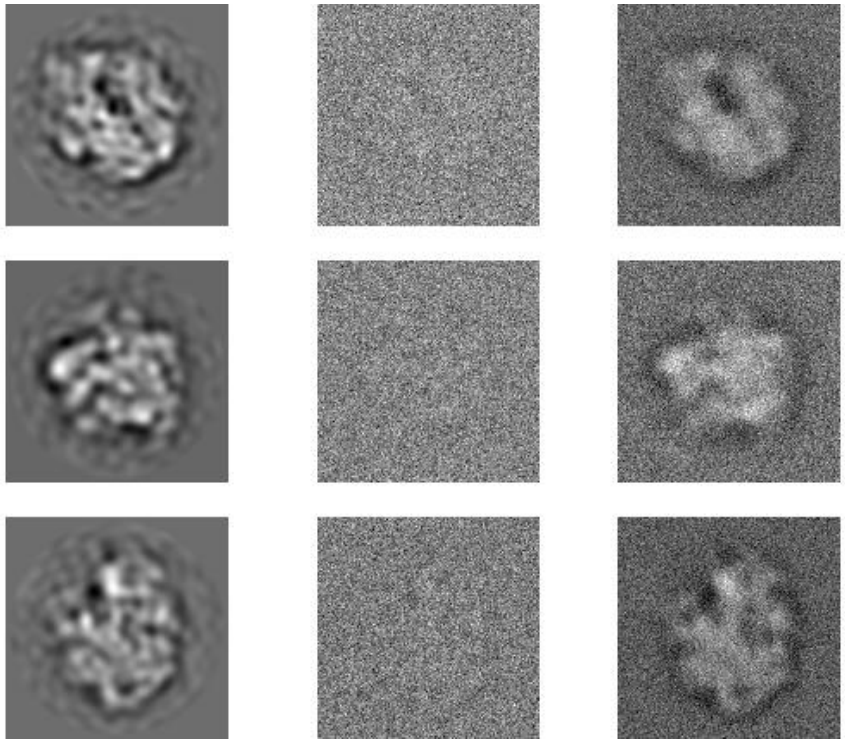


(a) Projection images. (b) Defocused images. (c) Noisy, defocused images.

Figure 12.1: Experimental setup. Input to global projection is a collection of 2.5×10^4 images which are noisy, defocused, and non-centered. Representatives are given in the rightmost column.

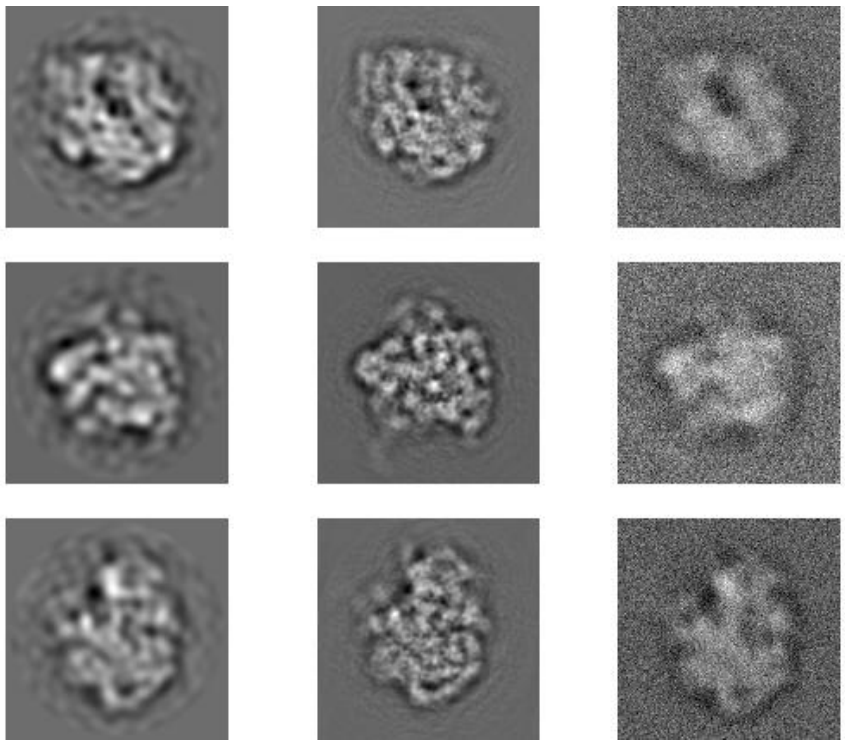
Class averaging and global projection were applied to the noisy image collection. Each of the class averages consisted of 200 images. Global pro-

jection denoised the Fourier coefficients of the images up to the band limit $b = 19.740\,\text{Å}^{-1}$, for each restriction of the class averaging operator 500 singular vectors were taken. Figures 12.2 and 12.3 compares the outputs of the two denoising algorithms.



(a) Global projection. (b) Noisy, defocused images. (c) Class averages.

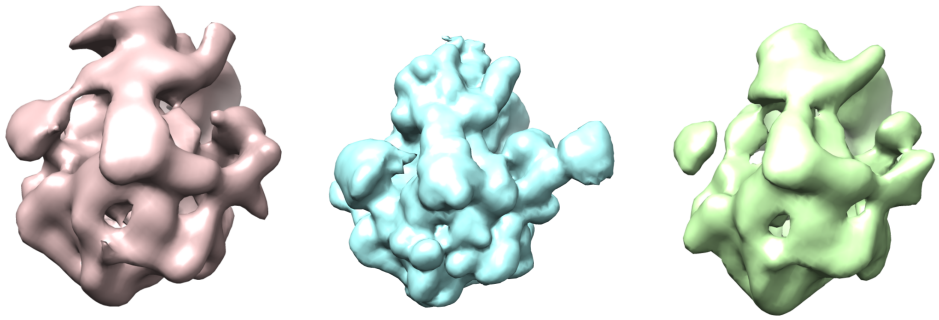
Figure 12.2: Comparison between global projection denoised images, noisy images, and class averages.



(a) Global projection. (b) Clean defocused images. (c) Class averages.

Figure 12.3: Comparison between global projection denoised images, clean defocused images, and class averages.

One thousand particles from the two denoised image collections were used to perform two separate reconstructions of the ribosome's potential function. The reconstructions were performed using the ASPIRE package. Specifically, least unsquared deviation was used to determine orientations [32] and an iterative linear solver was used to reconstruct the potential function [31].



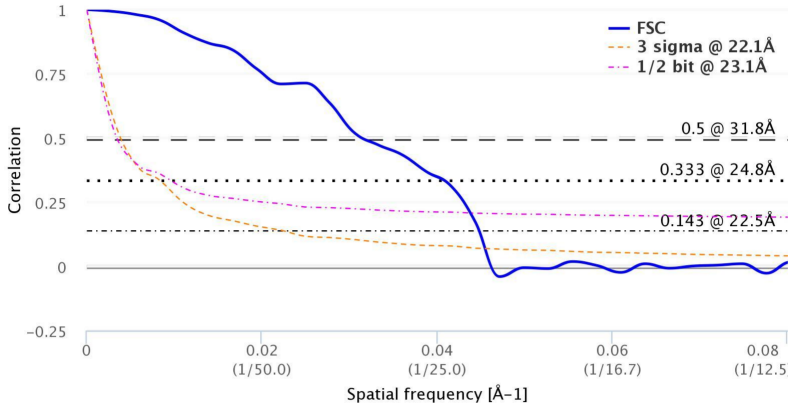
(a) Global projection. (b) Clean images. (c) Class averages.

Figure 12.4: Comparison between reconstruction using clean images and reconstructions using images denoised by global projection and class averaging. Potential functions visualized in Chimera. A Gaussian filter of width 1.5 was applied to all reconstructions, all potential functions visualized at cutoff 0.0473.

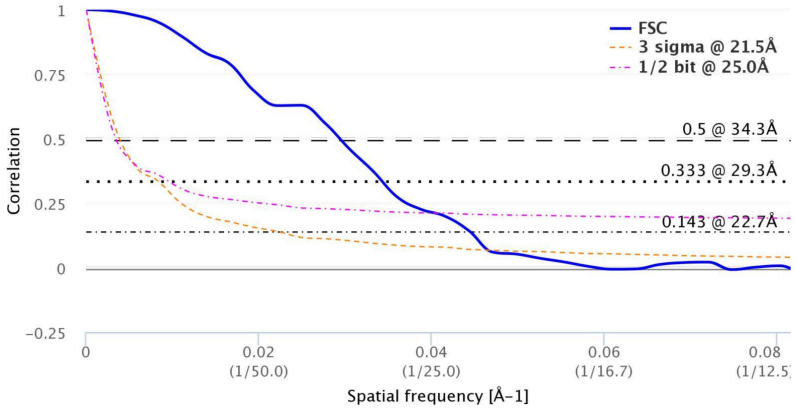
The quality of the reconstructions was assessed using the Fourier shell correlation (FSC) between the reconstructions and the true potential function [28]. Namely, For two potential function ϕ_1, ϕ_2 their FSC is defined as a function $\mathrm{FSC} : \mathbb{R}^+ \to \mathbb{C}$ with:

$$\mathrm{FSC}(r) = \frac{\int\limits_{v \in S_r} \widehat{\phi}_1(v)\widehat{\phi}_2(v)^* \, \mathrm{d}v}{\sqrt{\int\limits_{v \in S_r} |\widehat{\phi}_1(v)|^2 \, \mathrm{d}v \int\limits_{v \in S_r} |\widehat{\phi}_2(v)|^2 \, \mathrm{d}v}},$$

where S_r denotes the sphere of radius r and $\widehat{\phi}_1$, $\widehat{\phi}_2$ denotes the Fourier transform of ϕ_1 and ϕ_2 respectively. In words FSC takes as input a radius r and gives as output the correlation between the spatial frequency $1/r$ Fourier coefficients of ϕ_1 and ϕ_2.



(a) Global projection.



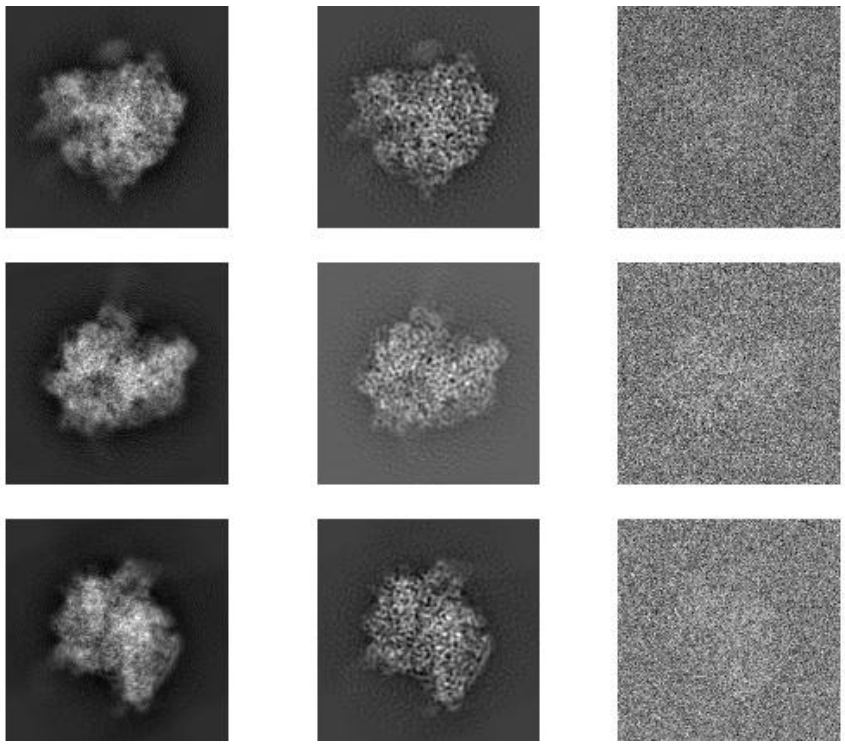
(b) Class averages.

Figure 12.5: FSC between the true potential function and reconstructions using images denoised by global projection and class averaging. FSC curves generated using the PDBe FSC server.

Using the 0.143 cutoff criteria the resolution of the class averaging reconstruction is 22.7 Å and the resolution of the global projection reconstruction is 22.5 Å.

12.1.2 Low contrast images

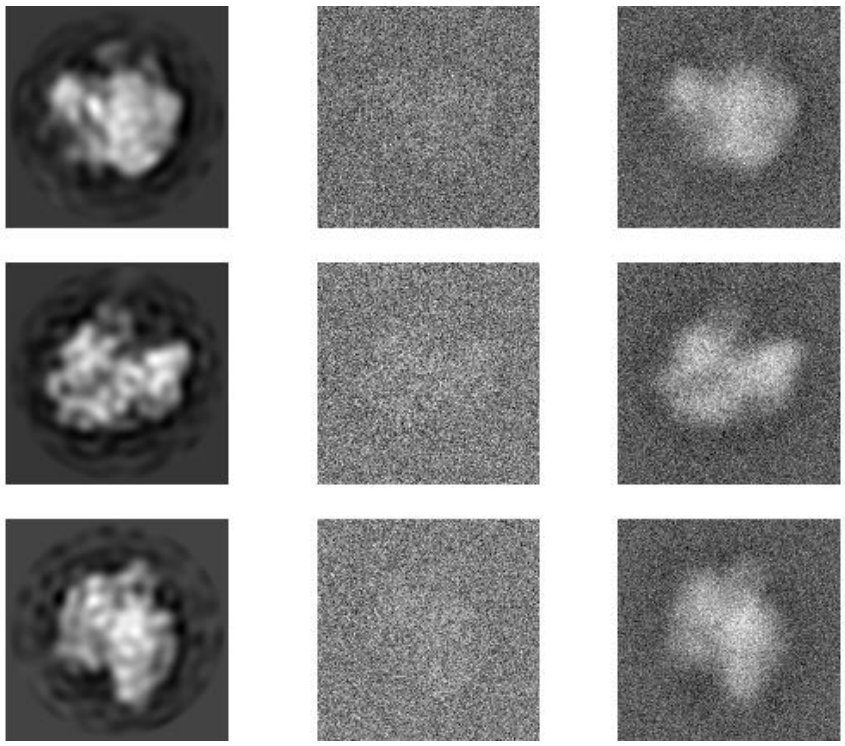
The ASPIRE package was used to generate 2.0×10^4 simulated images of a 50s ribosome subunit, each image had 129×129 pixels with a pixel length of 2.82 Å. The image collection consisted of a single defocus group with $\Delta z = 0.1\,\mu$m. White noise was added to give SNR $= 1/30$. Next, the images were phase flipped. Finally, the images were given a random shift between -5 to $+5$ pixels in both the x and y directions.



(a) Projection images. (b) Defocused images. (c) Noisy, defocused images.

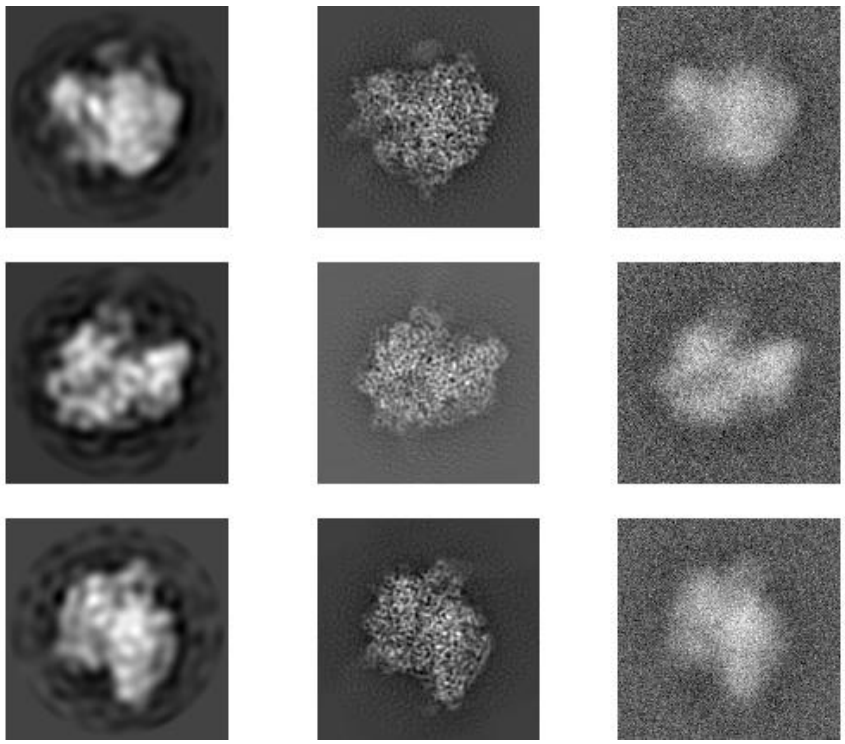
Figure 12.6: Experimental setup. Input to global projection is a collection of 2.0×10^4 images which are noisy, defocused, and non-centered. Representatives are given in the rightmost column.

Class averaging and global projection were applied to the noisy data set. Each of the class averages consisted of 100 images. Global projection denoised the Fourier coefficients of the images up to the band limit $b = 22.9239\,\text{Å}^{-1}$, for each restriction of the class averaging operator 500 singular vectors were taken. Figures 12.7 and 12.8 compares the outputs of the two denoising algorithms.



(a) Global projection. (b) Noisy, defocused images. (c) Class averages.

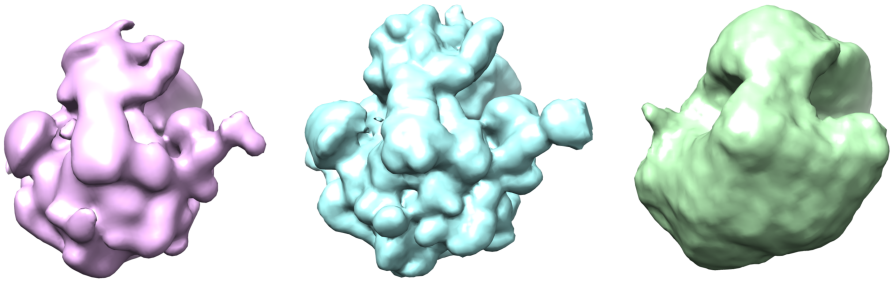
Figure 12.7: Comparison between global projection denoised images, noisy images, and class averages.



(a) Global projection. (b) Clean defocused images. (c) Class averages.

Figure 12.8: Comparison between global projection denoised images, clean defocused images, and class averages.

One thousand particles from the two denoised image collections were used to perform two separate reconstructions of the ribosome's potential function. The reconstruction were performed using the ASPIRE package.



(a) Global projection. (b) Clean images. (c) Class averages.

Figure 12.9: Comparison between reconstruction using clean images and reconstructions using images denoised by class averaging and global projection. Potential functions visualized in Chimera. A Gaussian filter of width 1.5 was applied to the reconstruction from true images and of 2.0 to the reconstruction from class averages.

Figures 12.10 and 12.11 report the FSC between the reconstructions and the true potential function. Using the 0.143 cutoff criteria the resolution of the class averaging reconstruction is 65.8 Å and the resolution of the global projection reconstruction is 26.1 Å.

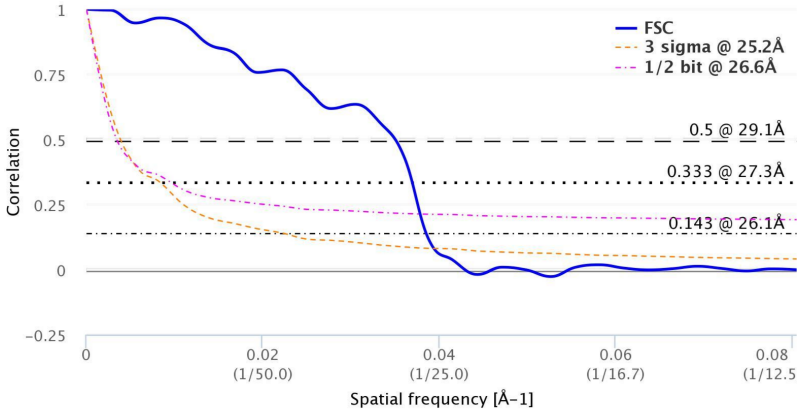


Figure 12.10: FSC between the true potential function and the reconstruction from images denoised by global projection.

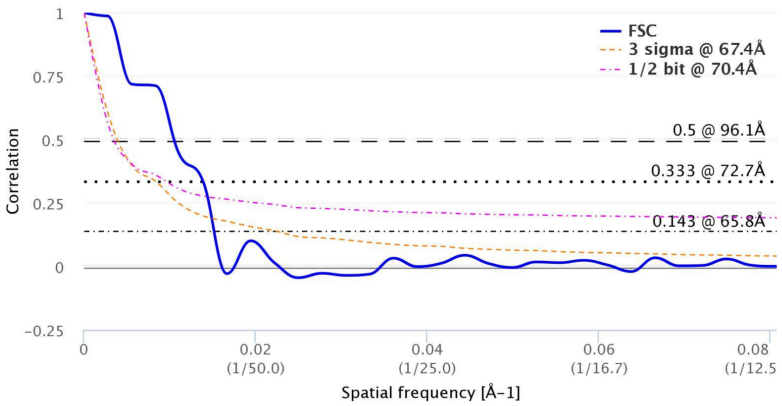


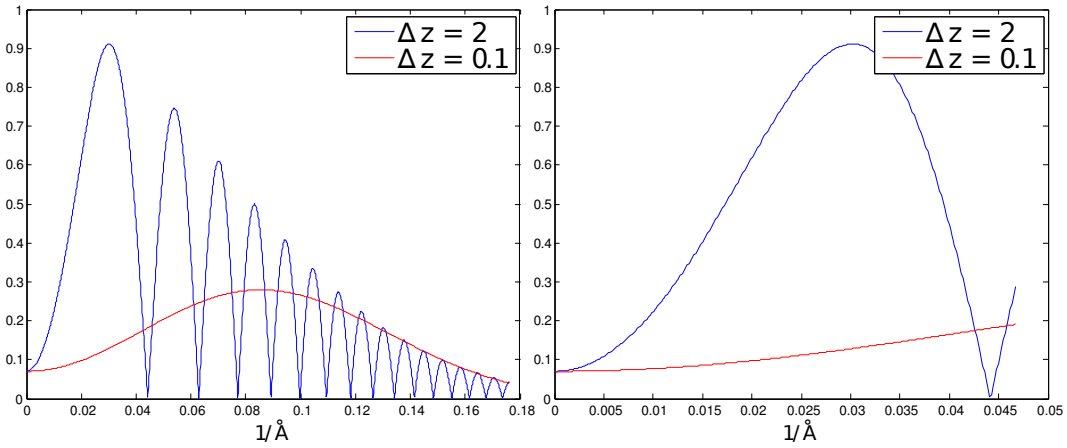
Figure 12.11: FSC between the true potential function and the reconstruction from class averages.

12.1.3 Discussion of results

We now discuss why global projection outperforms class averaging when applied to the low defocus images. Recall that phase flipping a Cryo-EM image taken from defcous Δz is equivalent to filtering the ideal projection image by the absolute value of the microscope's charge transfer function:

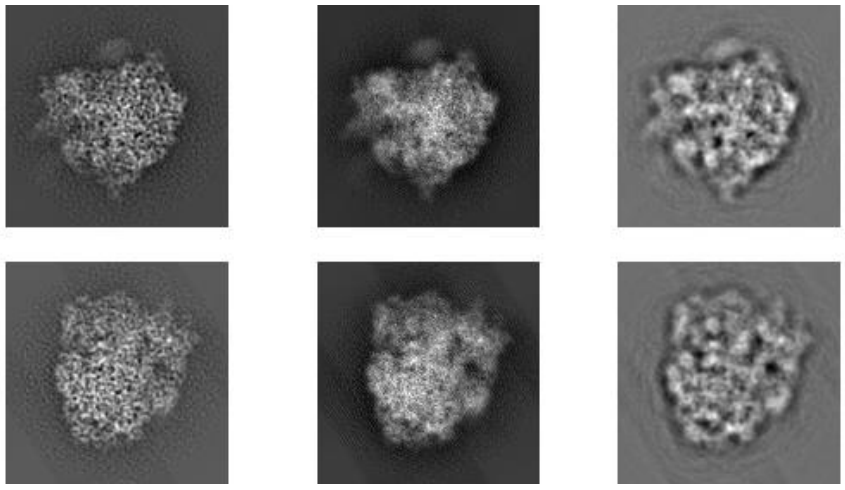
$$I_{flipped} = \mathcal{F}^{-1}\left(|\mathrm{CTF}_{\Delta z}|\mathcal{F}(I_{projection})\right),$$

where $\mathrm{CTF}_{\Delta z}$ denotes the charge transfer function. Figure 12.12 compares the absolute value of the CTFs used for the numerical experiment. We note that for small frequencies $\mathrm{CTF}_{0.1} \approx 0$, this means that the low defocused images are essentially low-pass filtered versions of the ideal projection images. Figure 12.13 compares the effect the different defocus values have on image formation.



(a) Comparision over all frequencies. (b) Comparision over low frequencies.

Figure 12.12: Comparision of the absolute value of the CTFs used in the numerical experiment.



(a) Defocus $\Delta z = 0.1$ (b) Ideal projection images. (c) Defocus $\Delta z = 2.0$

Figure 12.13: Comparison of defocused images to ideal projection images.

There are two reasons why the performance of class averaging degrades for low defocus images. First, class averaging requires groups of images to

be translationally aligned so that they can be averaged. Low defocus images have poor contrast which makes translationally aligning the images difficult. Second, the energy of low defocus images are concentrated in higher frequency Fourier coefficents which are more sensitive to translational shifts as seen in the well known identify:

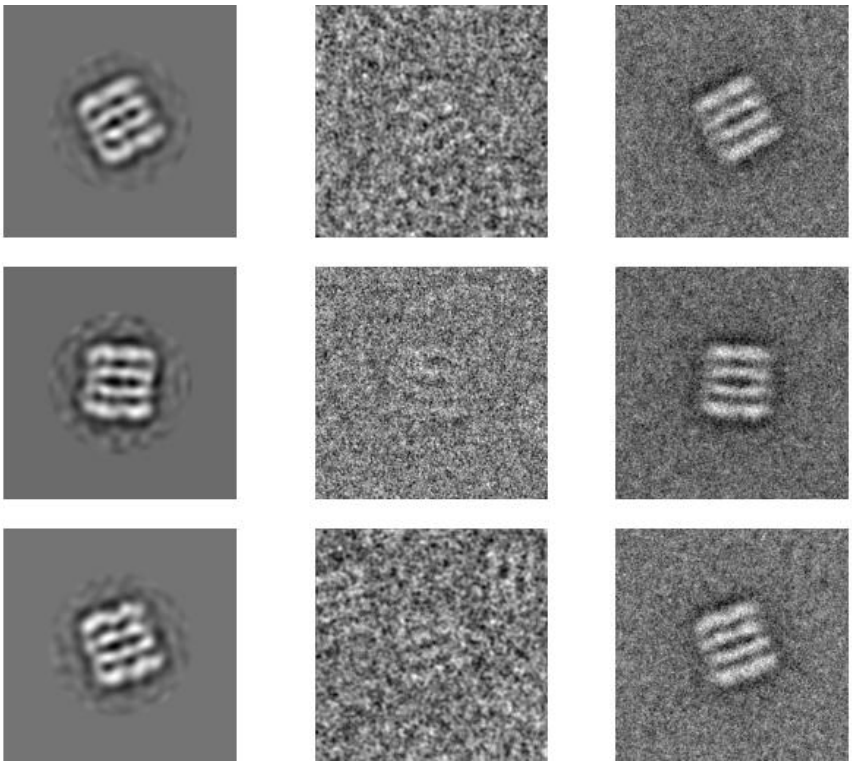
$$\mathcal{F}((s_x, s_y) \cdot I)(\xi_1, \xi_2) = \exp(2\pi i(s_x\xi_1 + s_y\xi_2))(\mathcal{F}I)(\xi_1, \xi_2),$$

for all shifts $(s_x, s_y) \in \mathbb{R}^2$ and frequencies $(\xi_1, \xi_2) \in \mathbb{R}^2$. Global projections remains effective when applied to low contrast images because the output images are centered.

12.2 Real data

12.2.1 GroEL

Global projection and class averaging were applied to a GroEL image set, publicly available at the Protein Data Bank in Europe [30]. The data set consisted of 2×10^4 images, each image had 159×159 pixels with a pixel length of 2.8 $\mathring{A}$. Global projection and class averaging were applied to the data. The class averages were of size 50. Global projection denoised the Fourier coefficients of the images up to the band limit $b = 21.5351\,\text{Å}^{-1}$, for each restriction of the class averaging operator 500 singular vectors were taken. The results is displayed in figure 12.14.



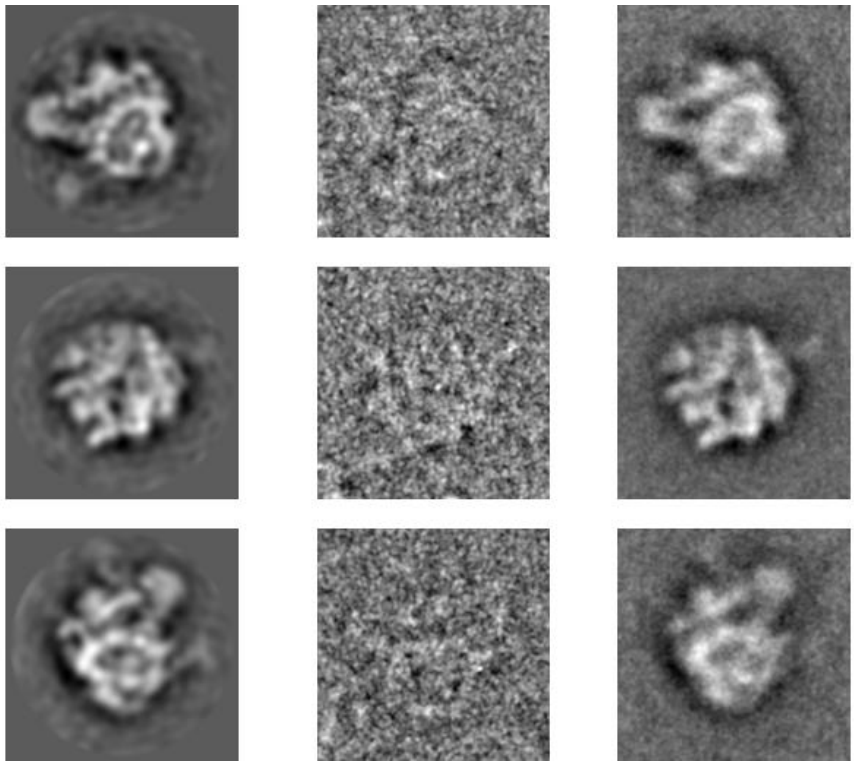
(a) Global projection. (b) Raw image. (c) Class average.

Figure 12.14: Comparision of global projection denoised images, raw images, and class averages.

12.2.2 50s ribosome subunit

Global projection and class averaging were applied to a ribosome image set, publicly available at the National Resource for Automated Molecular Microscopy website [30]. The data set consisted of 2×10^4 images, each image had 159×159 pixels with a pixel length of $1.63\,\text{Å}$. Global projection and class averaging were applied to the data. The class averages were of size 100. Global projection denoised the Fourier coefficients of the images up to the band limit $b = 18.9509\,\text{Å}^{-1}$, for each restriction of the class averaging operator

500 singular vectors were taken. The results is displayed in figure 12.15.



(a) Global projection. (b) Raw image. (c) Class average.

Figure 12.15: Comparision of global projection denoised images, raw images, and class averages.

Chapter 13

Conclusion

In the second part of this thesis we considered the problem of denoising collections of Cryo-EM images. First, we interpreted Cryo-EM image collections as a vector in a high dimensional space. Next, motivated by this interpretation we presented the global projection algorithm which works by projecting the image collection vector onto a low dimensional subspace. We showed that the output images are centered, and that their SNR grows linearly with the number of input images. The significance of the fact that the output images from global projection are centered was demonstrated by a numerical experiment where global projection successfully denoised a collection of simulated low contrast images.

13.1 Future work

A future line of research is to apply the global projection algorithm to the problem of aligning direct detector movie frames. Below we give an overview of the alignment problem and how global projection might be used to solve it.

13.1.1 Direct Detectors

Currently, Cryo-EM images are usually recorded indirectly using scintillators combined with CCD cameras (electrons excite photons through the scintillator which are then detected by the CCD camera). It takes a few seconds to record an image using this method, during this time two things happen which degrade image resolution. First, the electron beam causes the ice in which the specimen is embedded to move, thus blurring the recorded image. Second, the electron beam causes radiation damage to the specimen it is trying to image [1].

To overcome these two limitations a new generation of direct detectors have been introduced, which directly detect electrons in the silicon without using a scintillator. This means that instead of recording a single long exposure image of the sample, direct detectors can record movies of the sample consisting of $10 - 40$ frames. These movies can be used to correct for resolution loss due to sample movement and radiation damage. First, by aligning the frames of the movie and averaging, the sample movement can be corrected thus removing blurring due to sample movement. Second, studies have shown that radiation damage is frequency dependent, that is, radiation damages the higher frequencies of an image more then the lower frequencies. Therefore, by performing a frequency weighted averaging of the frames, radiation damage can be mitigated [34].

The problem of aligning Cryo-EM movie frames is difficult for two reasons. First, because each frame is the result of a very short electron beam exposure time, they have very low SNR. Second, the ice movement is local, that

is, different parts of the ice move differently. The following figure shows the particle movement over one second in a micrograph consisting of 70S ribosome subunits. The green dots are the starting positions of particles, and the red dots are their ending positions. The relative distances are exaggerated by a factor of 25 to improve clarity. One can see that nearby particles experience similar movements, but that far away particles can experience very different movements. The figure was taken from "Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles" [1].

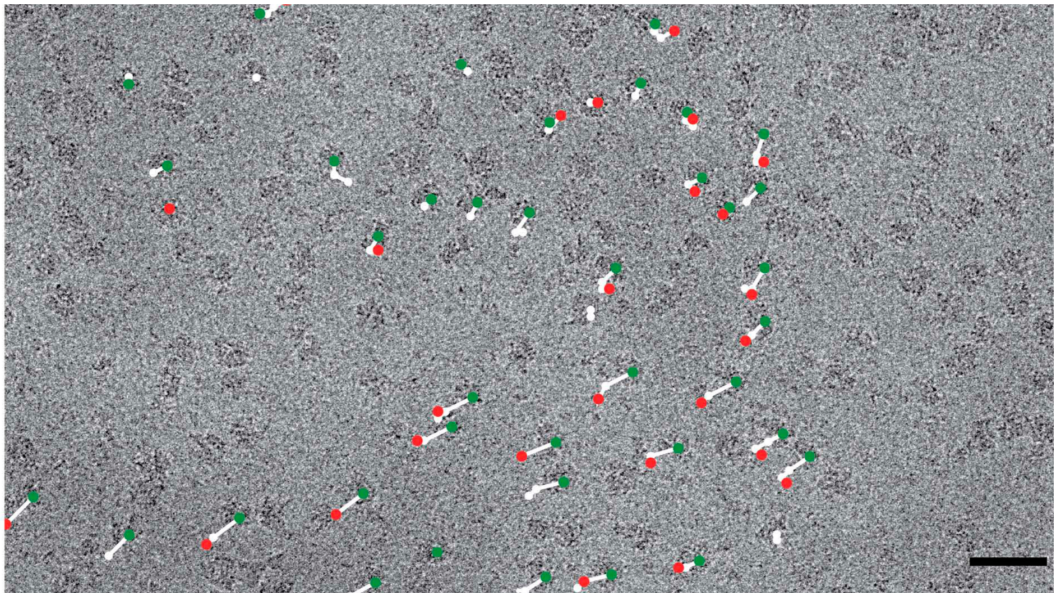


Figure 13.1: Movement of a 70S ribosome subunit micrograph.

Current algorithms align two movie frames by separately aligning different portions of the micrograph [15], [2]. The larger the size of the portions the more accurate the alignment calculations, but then the locality of the movements is not well represented.

In summary, when aligning Cryo-EM movie frames there is a trade-off between accuracy and locality. We note that each particle experiences a unique movement, therefore, to fully deblur the particle images the micrograph portions used for alignment should only contain a single particle. Any loss of locality limits the resolution of the final reconstruction.

13.1.2 Aligning direct detector movies with global projection

For direct detectors we have as input a collection of particle movies, $M_1, M_2, ..., M_n$. Where each movie records the movement of a single particle, and consists of a finite number of frames (typically 10-40):

$$M_i = \{F_{i,1}, ..., F_{i,m}\},$$

for all $i = 1, ..., n$ where m denotes the number of movie frames. By aligning the frames of each movie and averaging we can construct high SNR images of the particles. In this subsection we show how global projection can be used to do this.

The frames of each particle movie are translations of a single particle image. Therefore, the self correlation (or power spectrums) of the frames are the same up to noise, so by averaging them a good approximation of the particle image's self correlation (or power spectrum) can be computed. Using these approximations an emperical class averaging operator can be constructed. Taking the span of the top eigenvectors of this operator gives a vector space, denoted W, which well contains sections of clean projection images. We now consider the section of images corresponding to the first frame of the particle movies. The particles which produced these images had only been exposed to the electron

beam for a short time, so they contain high resolution details. Projecting this section onto W produces a section of clean centered images. The movie frames can then be aligned against these references, particle by particle. We note that in practice depending on the number of images and the SNR of the movie frames it may be necessary to construct a reference by projecting a section corresponding to an average of the first few movie frames onto W.

Appendices

Appendix A

Duality of finite sets and finite algebras

There exits a standard correspondence between finite sets and finite algebras. The correspondence takes the form of a dictionary $\mathcal{D}$ mapping finite sets to finite algebras and finite set maps to finite algebra morphisms.

$$\mathcal{D} : \text{Finite sets} \to \text{Finite algebras}$$

$$\mathcal{D} : \text{Finite set maps} \to \text{Finite algebra morphisms}$$

First, we define the correspondence between finite sets and finite algebras. Let S be a finite set, then $\mathcal{D}(S)$ is defined to be $\mathbb{C}[S]$, the complex valued functions on S. Next, we define the correspondence between finite set morphisms and finite algebra morphisms. Let $S1$ and $S2$ be finite sets and $s : S1 \to S2$ be a map between them. Then $\mathcal{D}(s)$ is defined to be the algebra morphism $\mathcal{D}(s) : \mathbb{C}[S2] \to \mathbb{C}[S1]$ given by:

$$(\mathcal{D}s)(f) = f \circ s,$$

for all $f \in \mathbb{C}[S2]$. We shall abbreviate $\mathcal{D}(s)$ by s^*, and call the processes of transforming s to $\mathcal{D}(s)$ as "taking the pullback of s".

Likewise there is a dual dictionary $\mathcal{D}'$ mapping finite algebras to finite sets and finite algebra morphisms to finite set maps.

$$\mathcal{D}' : \text{Finite algebras} \to \text{Finite sets}$$

$$\mathcal{D}' : \text{Finite algebra morphisms} \to \text{Finite set maps}$$

First we define the correspondence between finite algebras and finite sets. Let A be a finite algebra, then $\mathcal{D}'(A)$ is defined to be the spectrum of A, denoted by $\text{spec}(A)$. The spectrum of A consists of the non-zero complex characters of A. Recall complex characters are maps $\mathcal{X} : A \to \mathbb{C}$ which preserves multiplication i.e, $\mathcal{X}(ab) = \mathcal{X}(a)\mathcal{X}(b)$ for all $a, b \in A$. Next, we define the correspondence between finite algebra morphisms and finite set maps. Let $A1$ and $A2$ be finite algebras and $\alpha : A1 \to A2$ an algebra morphism between them. Then $\mathcal{D}'(\alpha)$ is defined to be a map $\mathcal{D}'(\alpha) : \text{spec}(A2) \to \text{spec}(A1)$ given by:

$$(\mathcal{D}'\alpha)(\mathcal{X}) = \mathcal{X} \circ \alpha,$$

for all $\mathcal{X} \in \text{spec}(A2)$. We shall abbreviate $\mathcal{D}'(\alpha)$ by α^*, and call the processes of transforming α to $\mathcal{D}'(\alpha)$ as "taking the pullback of α".

We now address the relationship between the dictionaries $\mathcal{D}$ and $\mathcal{D}'$. Note the composition $\mathcal{D}' \circ \mathcal{D}$ takes finite sets to finite sets and finite set maps to finite set maps. Ideally the composition would be the identity, but this is unfortunately not the case. Instead, the composition is equivalent to the identity by the canonical evaluation map. Recall for a finite set S the evaluation map, $ev : S \to \text{spec}(\mathbb{C}[S])$ is defined by: $[ev(x)](f) = f(x)$, for all $x \in S$ and $f \in \mathbb{C}[S]$. The following proposition characterizes how $\mathcal{D}' \circ \mathcal{D}$ is equivalent to the identity by the evaluation map.

Proposition 2. *For all pairs of finite sets $S1$, $S2$ and for all finite maps s with $s : S1 \to S2$. The following diagram commutes,*

$$\begin{array}{ccc} \text{spec}(\mathbb{C}[S1]) & \xrightarrow{s^{**}} & \text{spec}(\mathbb{C}[S2]) \\ \uparrow{\scriptstyle \text{ev}} & & \uparrow{\scriptstyle \text{ev}} \\ S1 & \xrightarrow{s} & S2 \end{array}$$

Furthermore the evaluation maps are bijections.

Proof. To see that the diagram commutes, note:

$$\begin{aligned}[s^{**}\mathrm{ev}(x)](f) &= [\mathrm{ev}(x)](s^*f) \\ &= (s^*f)(x) \\ &= f(s(x)) \\ &= [\mathrm{ev}(s(x))](f)\end{aligned}$$

for all $x \in S1$ and $f \in \mathbb{C}[S1]$. Clearly the evaluation maps are injective. To see surjectivity define the characteristic functions, $\mathbb{1}_x \in \mathbb{C}[S1]$ for all $x \in S1$, which is one at x and zero elsewhere. Then for all characters $\mathcal{X} \in \mathrm{spec}(\mathbb{C}[S1])$ note that $\mathcal{X}(\mathbb{1}_x) = \mathcal{X}(\mathbb{1}_x \cdot \mathbb{1}_x) = (\mathcal{X}(\mathbb{1}_x))^2$. Therefore $\mathcal{X}(\mathbb{1}_x)$ is either one or zero. Since $\mathcal{X}$ is non-zero and the characteristic functions provide a basis for $\mathbb{C}[S1]$ there most be some $a \in S1$ with $\mathcal{X}(\mathbb{1}_a) = 1$. Then, $0 = \mathcal{X}(0) = \mathcal{X}(\mathbb{1}_a \cdot \mathbb{1}_b) = \mathcal{X}(\mathbb{1}_a)\mathcal{X}(\mathbb{1}_b) = \mathcal{X}(\mathbb{1}_b)$ for all $b \neq a$. Putting the previous two calculations together gives,

$$\mathcal{X}(\mathbb{1}_x) = \begin{cases} 1 & \text{if} \quad x = a \\ 0 & \text{if} \quad x \neq a \end{cases}$$

The only way this can be true is if $\mathcal{X} = \mathrm{ev}(a)$. We conclude that the evaluation map is a bijection. □

An important property of the duality relationships is the functorality of the pullbacks, which says for maps s, t that $(s \circ t)^* = t^* \circ s^*$.

Proposition 3. ***(Functorality of pullbacks)***
For finite sets $S1, S2, S3$ and maps:

$$s : S1 \to S2$$

$$t : S2 \to S3,$$

we have that $(s \circ t)^* = t^* \circ s^*$.

Proof. The proof follows from applying the pullback definition:

$$\begin{aligned}(s \circ t)^*(f) &= f \circ s \circ t \\ &= (s^* f) \circ t \\ &= t^* \circ s^*(f),\end{aligned}$$

for all $f \in \mathbb{C}[S3]$. □

The functorality also holds for the correspondence of finite algebras to finite sets. That is, if α and β are algebra morphisms such that $\alpha \circ \beta$ is well defined, then: $(\alpha \circ \beta)^* = \beta^* \circ \alpha^*$.

In summary: the dictionaries $\mathcal{D}$, $\mathcal{D}'$ give a correspondence between finite sets and finite algebras. Proposition 2 shows that $\mathcal{D}' \circ \mathcal{D}$ is canonically equivalent by the evaluation map to the identity. It should be noted that $\mathcal{D} \circ \mathcal{D}'$ is also canonically equivalent to the identity, but this is not needed in this thesis and so will not be proved. Throughout this thesis we will write $\mathcal{D}' \circ \mathcal{D}(S) = S$ and $\mathcal{D}' \circ \mathcal{D}(s) = s$ with the understanding that equality is only true up to the evaluation map. In the thesis we will call the application of $\mathcal{D}$ and $\mathcal{D}'$ "taking the dual".

Appendix B

Proofs for part I

Proof of theorem 1. $\forall f \in \mathbb{C}[D]$:

$$\begin{aligned}(\mathcal{G}f)(d) &= \sum_{d' \in D} \delta_c(d, d') f(d') \\ &= \sum_{d' \in c^{-1}(c(d))} f(d') \\ &= [c^*(\mathcal{E}f)](d)\end{aligned}$$

Where $\mathcal{E} : \mathbb{C}[D] \to \mathbb{C}[C]$ with $(\mathcal{E}f)(p) = \sum_{d' \in c^{-1}(p)} f(d')$. This gives $\mathrm{Im}(\mathcal{G}) \subset c^*\mathbb{C}[C]$, for equality note $\mathcal{E}$ is surjective. 2. follows from 1. and the definition of $\mathcal{G}$.

□

Proof of theorem 2. Note that $\mathbb{E}\widehat{\mathcal{G}}$ decomposes as $\mathbb{E}\widehat{\mathcal{G}} = (1-q-p)\mathcal{G} + p\mathcal{A}$ with $\mathcal{A} : \mathbb{C}[D] \to \mathbb{C}[D]$ the averaging operator defined by,

$$(\mathcal{A}f)(d) = \sum_{d' \in D} f(d'),$$

for all $d \in D$ and $f \in \mathbb{C}[D]$. We note $\mathrm{Im}(\mathcal{A}) = \mathrm{span}(\{1_D\}) = c^*(\mathrm{span}(\{1_C\}))$, and by theorem 1 $\mathrm{Im}(\mathcal{G}) = c^*\mathbb{C}[C]$. Thus $\mathrm{Im}(\mathbb{E}\widehat{\mathcal{G}}) = c^*\mathbb{C}[C]$.

□

Proof of theorem 3. This will be done by relating the spectrum of $\mathcal{G}_r$ to the spectrum of $\mathcal{I} \otimes \mathcal{C}_r : \mathbb{C}[C] \otimes \mathbb{C}[X] \to \mathbb{C}[C] \otimes \mathbb{C}[X]$. Where $\mathcal{I} : \mathbb{C}[C] \to \mathbb{C}[C]$ is the identity operator. Note, $\mathcal{I} \otimes \mathcal{C}_r$ has eigenvalues λ_n for $n = 1, ..., |X|$ with corresponding eigenspaces $\mathbb{C}[C] \otimes V_n$. Define the multiplication map $m : \mathbb{C}[C] \otimes \mathbb{C}[X] \to \mathbb{C}[D]$ by:

$$(m(g \otimes f))(d) = g(c(d))f(x(d)),$$

for all $d \in D$ and $g \otimes f \in \mathbb{C}[C] \otimes \mathbb{C}[X]$. Then showing the following diagram commutes gives the result.

$$\begin{array}{ccc} \mathbb{C}[C] \otimes \mathbb{C}[X] & \xrightarrow{\mathcal{I} \otimes \mathcal{C}_r} & \mathbb{C}[C] \otimes \mathbb{C}[X] \\ \downarrow{\scriptstyle m} & & \downarrow{\scriptstyle m} \\ \mathbb{C}[D] & \xrightarrow{\mathcal{G}_r} & \mathbb{C}[D] \end{array}$$

The commutivity of the diagram is shown by applying the definitions. For all $g \otimes f \in \mathbb{C}[C] \otimes \mathbb{C}[X]$ we have:

$$\begin{aligned} [\mathcal{G}_r \circ m(g \otimes f)](d) &= \frac{1}{|B_r(x(d))|} \sum_{d' \in N_r(d)} \delta_c(d, d')[m(f \otimes g)](d') \\ &= \frac{1}{|B_r(x(d))|} \sum_{d' \in N_r(d)} \delta_c(d, d') f(x(d'))g(c(d')) \\ &= g(c(d)) \frac{1}{|B_r(x(d))|} \sum_{y \in B_r(x(d))} f(y) \\ &= (\mathcal{I}g)(c(d))(\mathcal{C}_r f)(x(d)) \\ &= [m(\mathcal{I} \otimes \mathcal{C}_r)(f \otimes g)](d). \end{aligned}$$

□

Proof of theorem 4. $\mathbb{E}\widehat{\mathcal{G}}_r$ decomposes as $\mathbb{E}\widehat{\mathcal{G}}_r = (1 - q - p)\mathcal{G}_r + p\mathcal{A}_r$. With

$\mathcal{A}_r \in \mathrm{End}(\mathbb{C}[D])$ the local averaging operator defined by,

$$(\mathcal{A}_r f)(d) = \frac{1}{|B_r(x(d))|} \sum_{d' \in N_r(d)} f(d'),$$

for all $d \in D$ and $f \in \mathbb{C}[D]$. The spectrum of $\mathbb{E}\widehat{\mathcal{G}}_r$ will be related to the spectrum of $((1-q-p)\mathcal{I}+p\mathcal{A})\otimes\mathcal{C}_r \in \mathrm{End}(\mathbb{C}[C]\otimes\mathbb{C}[X])$. With $\mathcal{A} : \mathbb{C}[C] \to \mathbb{C}[C]$ the averaging operator defined by,

$$(\mathcal{A} f)(b) = \sum_{a \in C} f(a),$$

for all $b \in C$ and $f \in \mathbb{C}[C]$. We note that $(1-q-p)\mathcal{I} + p\mathcal{A}$ has eigenvalues $(|C|p + (1-q-p)), (1-q-p)$ with corresponding eigenspaces $\mathbb{C}[C]_0, \mathbb{C}[C]_0^{\perp}$ respectivally. Therefore $\mathbb{C}[C]_0 \otimes \mathbb{C}[X]$ and $\mathbb{C}[C]_0^{\perp} \otimes \mathbb{C}[X]$ are invarient under $((1-q-p)\mathcal{I}+p\mathcal{A})\otimes\mathcal{C}_r$. On $\mathbb{C}[C]_0 \otimes \mathbb{C}[X]$ the operator has eigenvalues $[|C|p+(1-q-p)]\lambda_n$ for $n = 1, ..., |X|$ with corresponding eigenspaces $\mathbb{C}[C]_0 \otimes_m V_n$. On $\mathbb{C}[C]_0^{\perp} \otimes \mathbb{C}[X]$ the operator has eigenvalues $(1-q-p)\lambda_n$ for $n = 1, ..., |X|$ with corresponding eigenspaces $\mathbb{C}[C]_0^{\perp} \otimes V_n$. The result follows by noting:

$$m \circ (((1-q-p)\mathcal{I} + p\mathcal{A}) \otimes \mathcal{C}_r) = (\mathbb{E}\widehat{\mathcal{G}}_r) \circ m.$$

□

Proof of Proposition 1. First suppose a section s satisfies equation (4.5), then:

$$c_U^* \circ (c_W^*)^{-1} s(W) = t(W,U)^* s(W) = s(U),$$

for all $U, W \in \mathcal{U}$. Therefore, $(c_W^*)^{-1} s(W) = (c_U^*)^{-1} s(U)$ for all $U, W \in \mathcal{U}$. Let $f \in \mathbb{C}[C]$ be this common value, then $\tau(f) = s$ and so $s \in F$. For

the converse, suppose a section s is an element of F, then $s = \tau f$ for some $f \in \mathbb{C}[C]$, therefore:

$$\begin{aligned} t(W,U)^* s(W) &= c_U^* (c_W^*)^{-1} c_W^* f \\ &= c_U^* f \\ &= (\tau f)(U) \\ &= s(U) \end{aligned}$$

for all $U, W \in \mathcal{U}$. □

Proof of theorem 5. The first part follows from the fact that $\mathcal{C}$ is an ergodic markov chain (this is because $\mathcal{U}$ is connected under $\text{adj}_{\mathcal{U}}$). The proof of the second part will be done by relating the spectrum of $\mathcal{T}$ to the spectrum of $\mathcal{I} \otimes \mathcal{C} : \mathbb{C}[C] \otimes \mathbb{C}[\mathcal{U}] \to \mathbb{C}[C] \otimes \mathbb{C}[\mathcal{U}]$. Note $\mathcal{I} \otimes \mathcal{C}$ has eigenvalues λ_n for $n = 1, ..., |\mathcal{U}|$ with corresponding eigenspaces $\mathbb{C}[C] \otimes V_n$. Define a multiplication map $m_\tau : \mathbb{C}[C] \otimes \mathbb{C}[\mathcal{U}] \to \mathcal{H}$ by:

$$m_\tau(g \otimes f)(U) = c_U^* g f(U),$$

for all $U \in \mathcal{U}$ and $g \otimes f \in \mathbb{C}[C] \otimes \mathbb{C}[\mathcal{U}]$. The result follows by showing the following diagram commutes.

$$\begin{array}{ccc} \mathbb{C}[C] \otimes \mathbb{C}[\mathcal{U}] & \xrightarrow{\mathcal{I} \otimes \mathcal{C}} & \mathbb{C}[C] \otimes \mathbb{C}[\mathcal{U}] \\ \Big\downarrow{\scriptstyle m_\tau} & & \Big\downarrow{\scriptstyle m_\tau} \\ \mathcal{H} & \xrightarrow{\mathcal{T}} & \mathcal{H} \end{array}$$

Commutivity is shown by applying the definitions. For all $g \otimes f \in \mathbb{C}[C] \otimes \mathbb{C}[\mathcal{U}]$

we have:

$$\begin{aligned}
[(\mathfrak{T}m_\tau)(g \otimes f)](U) &= \frac{1}{|N(U)|} \sum_{W \in N(U))} c_U^*(c_W^*)^{-1} m_\tau(g \otimes f)(W) \\
&= \frac{1}{|N(U)|} \sum_{W \in N(U)} c_U^*(c_W^*)^{-1}(c_W^* g) f(W) \\
&= c_U^* g \frac{1}{|N(U)|} \sum_{W \in N(U)} f(W) \\
&= c_U^* g \, (\mathcal{C}f)(U) \\
&= [m_\tau(\mathcal{I} \otimes \mathcal{C})(g \otimes f)](U)
\end{aligned}$$

□

Proof of theorem 6. This will be done by relating the spectrum of $\mathbb{E}\widehat{\mathfrak{T}}$ to the spectrum of $(p\mathcal{I} + (1-p)P_0) \otimes \mathcal{C} : \mathbb{C}[C] \otimes \mathbb{C}[\mathcal{U}] \to \mathbb{C}[C] \otimes \mathbb{C}[\mathcal{U}]$. With P_0 the projection operator onto $\mathbb{C}[C]_0$. Note $p\mathcal{I} + (1-p)P_0$ has eigenvalues $1, p$ with corresponding eigenspaces $\mathbb{C}[C]_0, \mathbb{C}[C]_0^\perp$ respectivly. Therefore $(p\mathcal{I} + (1-p)P_0) \otimes \mathcal{C}$ has invarient subspaces $\mathbb{C}[C]_0 \otimes \mathbb{C}[\mathcal{U}]$ and $\mathbb{C}[C]_0^\perp \otimes \mathbb{C}[\mathcal{U}]$. On $\mathbb{C}[C]_0 \otimes \mathbb{C}[\mathcal{U}]$ the operator has eigenvalues λ_n for $n = 1, ..., |\mathcal{U}|$ with corresponding eigenspaces $\mathbb{C}[C]_0 \otimes V_n$. On $\mathbb{C}[C]_0^\perp \otimes \mathbb{C}[\mathcal{U}]$ the operator has eigenvalues $p\lambda_n$ for $n = 1, ..., |\mathcal{U}|$ with corresponding eigenspaces $\mathbb{C}[C]_0^\perp \otimes V_n$. The proposition follows by noting:

$$m_\tau \circ [(p\mathcal{I} + (1-p)P_0) \otimes \mathcal{C}] = \mathbb{E}\widehat{\mathfrak{T}} \circ m_\tau.$$

□

Appendix C

Proofs for part II

Proof of theorem 7. We first give a lower bound to the energy of the signal, $\iota^* P_{\mathcal{X}V_b} \mathcal{X}\phi$, where $P_{\mathcal{X}V_b}$ is orthoganal projection onto $\mathcal{X}V_b$.

$$\begin{aligned} energy(signal) &= ||\iota^* P_{\mathcal{X}V_b} \mathcal{X}\phi||^2 \\ &= \frac{1}{n} \sum_{\alpha=1}^{n} ||(P_{\mathcal{X}V_b} \mathcal{X}\phi)(\iota\alpha)||^2 . \end{aligned}$$

Let $E_{\phi,b}$ be the expected value of $||(P_{\mathcal{X}V_b} \mathcal{X}\phi)(\cdot)||^2$ under μ_o. We use Hoeffding's inequality [13] to prove that with exponential probability the energy of the signal is bounded below by $E_{\phi,b}/2$.

Theorem 14. *(**Hoeffding's inequality**) Let $X_1, X_2, ..., X_n$ be i.i.d instances of a random variable X. If $M > 0$ is a constant which bounds X almost surely, then:*

$$\mathbb{P}(|\overline{X}_n - \mathbb{E}[X]| \geq t) \leq 2\exp\left(\frac{-2nt^2}{M^2}\right),$$

where $\overline{X}_n$ is the emperical mean of the X_i's and $t > 0$.

We note that $||(P_{\mathcal{X}V_b} \mathcal{X}\phi)(\cdot)||^2$ is a continious function on the compact space $\mathcal{F}$, so has a finite upper bound $M_{\phi,b}$. Applying Hoeffding's inequality gives,

$$\Pr\left(energy(signal) \leq \frac{E_{\phi,b}}{2}\right) \leq 2\exp\left(\frac{-nE_{\phi,b}^2}{2M_{\phi,b}^2}\right). \tag{C.1}$$

We proceded to give an upper bound bound the energy of the noise. Recall, the section of observed images is $\iota^* P_{\mathcal{X}V_b}\mathcal{X}\phi + \epsilon$ where ϵ is the section of noise images. Therefore, the algorithm noise is given by:

$$\begin{aligned} noise &= projected\ section - signal \\ &= P_{\iota^*\mathcal{X}V_b}(\iota^*\mathcal{X}\phi + \epsilon) - \iota^* P_{\mathcal{X}V_b}\mathcal{X}\phi \\ &= P_{\iota^*\mathcal{X}V_b}(\iota^*(P_{\mathcal{X}V_b}\mathcal{X}\phi + P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi) + \epsilon) - \iota^* P_{\mathcal{X}V_b}\mathcal{X}\phi \\ &= P_{\iota^*\mathcal{X}V_b}(\iota^* P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi + \epsilon). \end{aligned}$$

To bound the energy of the noise it is nessasary to understand the projection operator $P_{\iota^*\mathcal{X}V_b}$. The least squares formula gives:

$$P_{\iota^*\mathcal{X}V_b} = \iota_b^*(\iota_b^{*\intercal}\iota_b^*)^{-1}\iota_b^{*\intercal},$$

where ι_b^* is ι^* restricted to $\mathcal{X}V_b$. The following lemma decompose $\iota_b^{*\intercal}$ and $(\iota_b^{*\intercal}\iota_b^*)^{-1}$ in terms of evaluation maps. To this end let $ev_\alpha : \mathcal{H} \to \mathcal{I}$ be defined by,

$$ev_\alpha(s) = s(\iota\alpha),$$

for $\alpha = 1, 2, ..., n$. Let ev_α^b be the restriction of ev_α to V_b.

Lemma 1. $\iota_b^{*\intercal}$ ***Decomposition***

$$\iota_b^{*\intercal} = \frac{1}{n}\sum_{\alpha=1}^{n} ev_\alpha^{b\intercal} ev_\alpha$$

$$\iota_b^{*\intercal}\iota_b^* = \frac{1}{n}\sum_{\alpha=1}^{n} ev_\alpha^{b\intercal} ev_\alpha^b$$

The Lemma's proof immeditally follows this proof. The energy of the noise can be bounded using the Cauchy-Schwartz inequality:

$$\begin{aligned} energy(noise) &= \left\langle P_{\iota^*\mathcal{X}V_b}(\iota^* P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi + \epsilon), \iota^* P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi + \epsilon \right\rangle \\ &= \left\langle (\iota_b^{*\intercal}\iota_b^*)^{-1}\iota_b^{*\intercal}(\iota^* P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi + \epsilon), \iota_b^{*\intercal}(\iota^* P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi + \epsilon) \right\rangle \quad \text{(C.2)} \\ &\leq \left|\left|(\iota_b^{*\intercal}\iota_b^*)^{-1}\right|\right| \left|\left|\iota_b^{*\intercal}(\iota^* P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi + \epsilon)\right|\right|^2. \end{aligned}$$

Inequality C.2 in turn will be given a exponentialy likely $^{\dim(V_n)}/_n$ bound in two steps. First $||(\iota_b^{*\intercal}\iota_b^*)^{-1}||$ will be given a exponetially likely bound using a matrix Chernoff bound [26]. Second $\iota_b^{*\intercal}(\iota^* P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi + \epsilon)$ will be shown to be mean zero, and it's squared norm will be given an exponetially likely $^{\dim(V_b)}/_n$ bound using Hoeffding's inequality. We state below the matrix Chernoff bound we use.

Theorem 15. ***(Matrix Chernoff bound)*** *Let $M_1, M_2, ..., M_n$ be i.i.d. instances of a random positive definite matrix M. Let $\mu_{\min}$ be the smallest eigenvalue of $\mathbb{E}M$. If $R > 0$ is a constant which bounds the spectrum of M almost surely, then:*

$$\Pr\left(\lambda_{\min}(\overline{M}_n) \leq \frac{\mu_{\min}}{4}\right) \leq d(.667)^{\frac{n\mu_{\min}}{R}},$$

where $\overline{M}_n$ is the emperical mean of the M_i's and d is the dimension of the matricies.

We note that the matricies $ev_{(\cdot)}^{b\intercal} ev_{(\cdot)}^b$ are positive definite. Taking the norm of these matricies gives a continious function on the compact space $\mathcal{F}$ which has a finite upper bound M_b. Finally a straghtforward calculation shows that the expected value of $ev_{(\cdot)}^{b\intercal} ev_{(\cdot)}^b$ is the identity operator. Using these observations we can apply the matrix Chernoff bound to the decomposition of $\iota_b^{*\intercal}\iota_b^*$ given in lemma 1 to get:

$$\Pr\left(\left|\left|(\iota_b^{*\intercal}\iota_b^*)^{-1}\right|\right| \geq 4\right) \leq \dim(V_b)(.667)^{\frac{n}{M_b}}. \tag{C.3}$$

We procede to bound the squared norm of the section $\iota_b^{*\intercal}(\iota^* P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi + \epsilon)$ using Hoeffding's inequality. To this end we apply lemma 1 to express the section as a sum of i.i.d. random vectors:

$$\iota_b^{*\intercal}(\eta) = \frac{1}{n}\sum_{\alpha=1}^{n} ev_\alpha^{b\intercal} ev_\alpha(\eta), \tag{C.4}$$

where η denotes $\iota^* P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi + \epsilon$. The vectors appearing on the right hand side are i.i.d. instances of the random vector $ev_{(\cdot)}^{b\intercal} ev_{(\cdot)}(\eta)$. The random vector is mean zero, i.e. the expected inner product between the random vector and $s \in \mathcal{X}V_b$ is zero. To see this note:

$$\mathbb{E}\left[\left\langle s, ev_\gamma^{b\intercal} ev_\gamma(\eta)\right\rangle\right] = \mathbb{E}\left[\left\langle s(\gamma), \left(P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi\right)(\gamma)\right\rangle\right] + \mathbb{E}\left[\langle s(\gamma), \epsilon(\gamma)\rangle\right],$$

where expectation is taken over the random orientation γ. The first term is clearly zero and the second term is zero by the assumptions placed on the noise. By continuity on the compact set $\mathcal{F}$ we have that $\left|\left|ev_{(\cdot)}^{b\intercal}\right|\right|$ has a finite upper bound C_1 and $\left|\left|P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi\right|\right|$ has a finite upper bound C_2. Thus we get the bound:

$$\left|\left|ev_\alpha^{b\intercal} ev_\alpha(\eta)\right|\right| = \left|\left|ev_\alpha^{b\intercal} ev_\alpha(P_{(\mathcal{X}V_b)^\perp}\mathcal{X}\phi + \epsilon)\right|\right| \leq C_1(C_2 + \sigma),$$

where we used the assumption that the norm of the noise images are bounded by σ. We will now use Hoeffding's inequality to bound the cordinates of the noise section with exponential probability, then use a union bound to give a bound on the squared norm of the error section. To this end note equation (C.4) gives:

$$\langle \iota_b^{*\intercal}\eta, s\rangle = \frac{1}{n}\sum_{\alpha=1}^{n}\left\langle ev_\alpha^{b\intercal} ev_\alpha(\eta), s\right\rangle,$$

for all $s \in \mathcal{X}V_b$ with $||s|| = 1$. The right hand side is the emperical mean of n i.i.d. instances of a random variable which is mean zero and absolute value bounded by $C_1(C_2 + \sigma)$. Thus Hoeffding's inequality gives:

$$\Pr\left(|\left\langle \iota_b^{*\intercal}\eta, s\right\rangle| \geq t\right) \leq 4\exp\left(\frac{-nt^2}{C_1^2(C_2+\sigma)^2}\right), \tag{C.5}$$

for all $t > 0$. Equation (C.5) can be applied to an orthonormal basis of $\mathcal{X}V_b$, applying a union bound to the result gives:

$$\Pr\left(\left|\left|\iota_b^{*\intercal}\eta\right|\right|^2 \geq t\dim(V_b)\right) \leq 4\dim(V_b)\exp\left(\frac{-nt}{C_1(C_2+\sigma)}\right), \tag{C.6}$$

for all $t > 0$. Combining equations (C.6) and (C.3) gives,

$$\Pr\left(energy(noise) \geq \frac{4\sigma^2 \dim(V_b)}{n\lambda}\right) \leq 4\dim(V_b)\exp\left(\frac{-\sigma^2}{\lambda C_1(C_2+\sigma)}\right)... \\ + \dim(V_b)(.667)^{\frac{n}{M_b}},$$

for all $\lambda > 0$. Combining this upper bound for the energy of the noise with the upper bound for the energy of the signal given in the equation (C.1) we get the desired bound on the SNR.

□

Proof of lemma 1. For all $s \in \widehat{\mathcal{H}}$ and $v \in V_b$ we have:

$$\begin{aligned}\langle \iota_b^{*\intercal} s, v\rangle &= \langle s, \iota_b^* v\rangle \\ &= \frac{1}{n}\sum_{\alpha=1}^{n} \langle s(\alpha), v(\alpha)\rangle \\ &= \left\langle \frac{1}{n}\sum_{\alpha=1}^{n} \mathrm{ev}_\alpha^{b\intercal}\mathrm{ev}_\alpha s, v\right\rangle.\end{aligned}$$

Therefore:

$$\iota_b^{*\intercal} = \frac{1}{n}\sum_{\alpha=1}^{n} \mathrm{ev}_\alpha^{b\intercal}\mathrm{ev}_\alpha.$$

Note $\mathrm{ev}_\alpha \circ \iota_b^* = \mathrm{ev}_\alpha^b$, so:

$$\iota_b^{*\intercal}\iota_b^* = \frac{1}{n}\sum_{\alpha=1}^{n} ev_\alpha^{b\intercal} ev_\alpha^b.$$

□

Proof of theorem 9. The proof works by showing eigensections s of Δ_{fiber} have the form:

$$s(x) = f(x)\psi^{k,q},$$

for all $x \in \mathcal{F}$, where $f \in \mathbb{C}[\mathcal{F}]$ and $\psi^{k,q}$ is an eigenfunction of $\Delta_{\mathbb{R}^2}$. To do this properly we need to understand the $SO(V)$ decomposition of $\mathbb{C}[\mathcal{F}]$, where $SO(V)$ acts on $\mathbb{C}[\mathcal{F}]$ by the left regular action, that is, $(g \cdot f)(x) = f(g^{-1}x)$ for all $f \in \mathbb{C}[\mathcal{F}]$ and $g \in SO(V)$. Additionally, let $SO(\mathbb{R}^2)$ act on $\mathbb{C}[\mathcal{F}]$ by the right regular action, that is, $(h \cdot f)(x) = f(xh)$ for all $f \in \mathbb{C}[\mathcal{F}]$ and $h \in SO(\mathbb{R}^2)$. The actions commute and so give an isotypic $SO(V)$ decomposition of $\mathbb{C}[\mathcal{F}]$, that is:

$$\mathbb{C}[\mathcal{F}] = \bigoplus_{l=0}^{\infty} \bigoplus_{k=-\infty}^{\infty} F_l^k. \tag{C.7}$$

Then for all $f \in F_l^k$ we have, $R_\theta \cdot g = e^{ik\theta} g$, where $R_\theta \in SO(\mathbb{R}^2)$ is rotation by θ and F_l^k is the sum of multiple copies of the l'th representation of $SO(V)$. The following theorem states that this decomposition is irreducable.

Theorem 16. *$SO(V)$ **decomposition of** $\mathbb{C}[\mathcal{F}]$*
Decomposition (C.7) is irreducable. Furthermore $F_l^k = 0$ for all $|k| > l$.

For the proof see [12].

Proof of theorem 9.1. Suppose $|k_b| > l$, we will show that non-zero sections of $E_l^{k_b,k_f,q}$ induce non-zero functions in $F_l^{k_b} \subset \mathbb{C}[\mathcal{F}]$ for all $k_f \in \mathbb{Z}$ and $q \in \mathbb{N}^+$. The result then follows from contradiction because theorem 16 states that $F_l^{k_b} = 0$.

Let $s \in E_l^{k_b,k_f,q}$ be a non-zero section. Then there exist a point $a \in \mathcal{F}$ with $s(a) \neq 0$. Consider the map $\alpha : E_l^{k_b,k_f,q} \to \mathbb{C}[\mathcal{F}]$ defined by:

$$(\alpha r)(x) = \langle r(x), s(a) \rangle .$$

Clearly α intertwines the $SO(V)$ actions. Also, α intertwines the π_{base} action on $E_l^{k_b,k_f,q}$ with the $SO(\mathbb{R}^2)$ action on $\mathbb{C}[\mathcal{F}]$. Therefore, by Schur's lemma we

have $\alpha E_l^{k_b,k_f,q} \subseteq F_l^{k_b}$. To see $\alpha(s) \neq 0$ note $(\alpha s)(a) = \langle s(a), s(a)\rangle = ||s(a)||^2 \neq 0$. $\square$

Proof of theorem 9.2 and 9.3. For all non-zero $s \in E_l^{k_b,k_f,q}$ where $|k_f| \leq l$, we know how the actions π_{fiber} and Δ_{fiber} affect s:

$$\Delta_{fiber} s = s\lambda_l^{k_b,k_f,q}$$
$$R_\theta \cdot s = e^{ik_f\theta} s,$$

where $R_\theta \in SO(\mathbb{R}^2)$ is rotation by θ. The fiberwise definitions of the actions means that:

$$\begin{aligned} \Delta_{\mathbb{R}^2} s(x) &= s(x)\lambda_l^{k_b,k_f,q} \\ s(x) \circ R_\theta &= e^{ik_f\theta} s(x), \end{aligned} \tag{C.8}$$

for all $x \in \mathcal{F}$. Choose some $a \in \mathcal{F}$ where $s(a) \neq 0$. Equation (C.8) shows $s(a)$ is an eigenfunction of $\Delta_{\mathbb{R}^2}$, so $\lambda_l^{k_b,k_f,q}$ must be the eigenvalue, $\mu^{k',q'}$, of $\Delta_{\mathbb{R}^2}$ for some unkown value of $q' \in \mathbb{N}^+, k' \in \mathbb{Z}$, with $s(a)$ a scaler multiple of the corresponding eigenfunction $\psi^{k',q'}$. In fact since equation (C.8) holds for all $x \in \mathcal{F}$ we have that s is pointwise a scaler multiple of the same eigenfunction $\psi^{k',q'}$. To narrow down the eigenfunction note the second part of equation (C.8) gives:

$$R_\theta \cdot s(a) = e^{ik_f\theta} s(a)$$

Therefore $s(a)$ must be proportional to $\psi^{k_f,q'}$ and $\lambda_l^{k_b,k_f,q} = \mu^{k_f,q'}$ for some unkown q'. The final step is to show $q' = q$. Since the eigenvalues of both $\Delta_{\mathbb{R}^2}$ and Δ_{fiber} increase in the $\mathbb{N}^+$ index showing $q = q'$ can be done by showing:

$$\left\{\mu^{k_f,n}\right\}_{n\in\mathbb{N}^+} = \left\{\lambda^{l,k_b,k_f,m}\right\}_{m\in\mathbb{N}^+}$$

The previous discussion gives $\supseteq$. To see $\subseteq$ we need to construct eigensections. Consider the map $\beta^{k_f,n} : \mathbb{C}[\mathcal{F}] \to \mathcal{H}$ defined by:

$$(\beta^{k_f,n} g)(x) = g(x)\psi^{k_f,n},$$

for all $n \in \mathbb{N}^+$. Note $\beta^{k_f,n}$ is injective, so by choosing any non-zero $g \in \mathbb{C}[\mathcal{F}]$ we get a non-zero section $\beta^{k_f,n} g \in \mathcal{H}$ which by construction is an eigenvector of Δ_{fiber} with eigenvalue $\mu^{k_f,m}$, thus:

$$\left\{\mu^{k_f,n}\right\}_{n\in\mathbb{N}^+} \subseteq \left\{\lambda^{l,k_b,k_f,m}\right\}_{m\in\mathbb{N}^+}$$

□

Proof of theorem 9.4. For each $k_f \in \mathbb{Z}, q \in \mathbb{N}^+$ consider two maps:

$$\alpha^{k_f,q} : E^{k_f,q} \to \mathbb{C}(\mathcal{F})$$
$$\beta^{k_f,q} : \mathbb{C}(\mathcal{F}) \to E^{k_f,q}.$$

Defined by:

$$(\alpha^{k_f,q} s)(x) = \left\langle s(x), \psi^{k_f,q} \right\rangle$$
$$(\beta^{k_b,q} g)(x) = g(x)\psi^{k_f,q},$$

Theorem 9.3 states that $\alpha^{k_f,q}$ and $\beta^{k_f,q}$ are inverses of each other. Note that the maps intertwine the $SO(V)$ actions and the $SO(\mathbb{R}^2)$ action on $\mathbb{C}[\mathcal{F}]$ with the π_{base} action on $E^{k_f,q}$. Therefore, by Schur's lemma,

$$\alpha^{k_f,q} E_l^{k_b,k_f,q} = F_l^{k_b}.$$

To complete the proof we invoke theorem 16 which states that $F_l^{k_b}$ is irreducable. □

□

Proof of theorem 10. We first prove that $\mathcal{X}$ is an $SO(V)$ intertwinner. This is

done using equation (8.1) which relates $\mathcal{X}$ to the standard X-ray transfrom:

$$\begin{aligned}\mathcal{X}(g\cdot\phi)(x) &= X_{std}(x^*(\phi\circ g^{-1}))\\ &= X_{std}(\phi\circ g^{-1}\circ x)\\ &= X_{std}(\phi\circ g^{-1}x)\\ &= X_{std}((g^{-1}x)^*\phi)\\ &= (g\cdot\mathcal{X}(\phi))(x),\end{aligned}$$

for all $g\in SO(V)$ and $\phi\in C_0(B_V)$. To prove the equivalence of the two $SO(\mathbb{R}^2)$ we again use equation (8.1):

$$\begin{aligned}(\pi_{base}(h)\mathcal{X}(\phi))(x) &= X_{std}((xh)^*\phi)\\ &= X_{std}(\phi\circ x\circ h)\\ &= h\cdot X_{std}(x^*\phi)\\ &= (\pi_{fiber}(h)\mathcal{X})(x),\end{aligned}$$

for all $h\in SO(\mathbb{R}^2)$ and $\phi\in C_0(B_V)$. □

Proof of theorem 12. To prove 12.1 recall that $(\pi_{fiber}(h)s)(x) = s(x)\circ h$, for all $s\in\mathcal{H}$ and $h\in SO(\mathbb{R}^2)$. Then applying the class averaging operator to the π_{fiber} action gives:

$$\begin{aligned}(\mathcal{C}\pi_{fiber}(h)s)(x) &= \frac{1}{|N_x|}\int\limits_{y\in N_x} k(x,y)\cdot(s(y)\circ h)\,\mathrm{d}\mu_0(y)\\ &= \left(\frac{1}{|N_x|}\int\limits_{y\in N_x} k(x,y)s(y)\,\mathrm{d}\mu_0(y)\right)\circ h\\ &= (\pi_{fiber}(h)\mathcal{C}s)(x).\end{aligned}$$

To prove 12.2 recall that $(\pi_{base}(h)s)(x) = s(xh)$, for all $s\in\mathcal{H}$ and $h\in SO(\mathbb{R}^2)$. We first note that teh class averaging kernel is $SO(\mathbb{R}^2)$ invarient in the sense

that $k(xh, xh) = k(x, y)$ for all $h \in SO(\mathbb{R}^2)$. The invarience follows directly from the definition of k:

$$\begin{aligned} k(xh, yh) &= \pi_{1,2} \circ h^{-1}x^{-1} \circ R_{x_3,y_3} \circ yh \circ \iota_{1,2} \\ &= h^{-1}\pi_{1,2} \circ x^{-1} \circ R_{x_3,y_3} \circ y \circ \iota_{1,2}h \\ &= \pi_{1,2} \circ x^{-1} \circ R_{x_3,y_3} \circ y \circ \iota_{1,2}h^{-1}h \\ &= k(x, y) \end{aligned}$$

The second equality follows from the fact that $\pi_{1,2}$ intertwines the $SO(\mathbb{R}^2)$ actions, the third equality follows from the fact that rotaions in the plane commute. Using the invarience of k we can prove that $\mathcal{C}$ intertwines the π_{base} action:

$$\begin{aligned} (\mathcal{C}\pi_{base}(h)s)(x) &= \frac{1}{|N_x|} \int\limits_{y \in N_x} k(x, y) \cdot s(yh) \, \mathrm{d}\mu_o(y) \\ &= \frac{1}{|N_x|} \int\limits_{y \in N_x} k(x, yh^{-1}) \cdot s(y) \, \mathrm{d}\mu_o(y) \\ &= \frac{1}{|N_{xh}|} \int\limits_{y \in N_{xh}} k(xh, y) \cdot s(y) \, \mathrm{d}\mu_o(y) \\ &= (\pi_{fiber}(h)\mathcal{C}s)(x) \end{aligned}$$

The second equality follows from the fact that μ_o is invarient under right multiplication of $SO(\mathbb{R}^2)$ and that $N_h(x) = N_h(xh)$. To prove 12.3, note that $k(x, y)h = k(x, yh)$ for all $h \in SO(\mathbb{R}^2)$. Then applying the class averaging

oeprator to the π_{fiber} action gives:

$$\begin{aligned}
(\mathcal{C}\pi_{fiber}(h)s)(x) &= \frac{1}{|N_x|}\int\limits_{y\in N_x} k(x,y)\cdot s(y)\circ h\,\mathrm{d}\mu_o(y)\\
&= \frac{1}{|N_x|}\int\limits_{y\in N_x} s(y)\circ k(x,y)^{-1}h\,\mathrm{d}\mu_o(y)\\
&= \frac{1}{|N_x|}\int\limits_{y\in N_x} s(y)\circ k(x,yh^{-1})^{-1}\,\mathrm{d}\mu_o(y)\\
&= \frac{1}{|N_x|}\int\limits_{y\in N_x} k(x,yh^{-1})\cdot s(y)\,\mathrm{d}\mu_o(y)\\
&= \frac{1}{|N_{xh}|}\int\limits_{y\in N_h(xh)} k(xh,y)\cdot s(y)\,\mathrm{d}\mu_o(y)\\
&= (\pi_{base}(h)\mathcal{C}s)(x).
\end{aligned}$$

The proof of 12.4 follows from the fact that the two dimensional Laplacian commutes with rotations:

$$\begin{aligned}
(\mathcal{C}\Delta_{fiber}s)(x) &= \frac{1}{|N_x|}\int\limits_{y\in N_x} k(x,y)\cdot \Delta_{\mathbb{R}^2}s(y)\,\mathrm{d}\mu_o(y)\\
&= \Delta_{\mathbb{R}^2}\frac{1}{|N_x|}\int\limits_{y\in N_x} k(x,y)\cdot s(y)\,\mathrm{d}\mu_o(y)\\
&= (\Delta_{fiber}\mathcal{C}s)(x)
\end{aligned}$$

The proof of 12.5 follows from the fact that k is $SO(V)$ invarient in the sense that $k(gx,gy)=k(x,y)$ for all $g\in SO(V)$. The invariance follows directly from the definition of k:

$$\begin{aligned}
k(gx,gy) &= \pi_{1,2}\circ x^{-1}g^{-1}\circ R_{gx_3,gy_3}\circ gy\circ\iota_{1,2}\\
&= \pi_{1,2}\circ x^{-1}g^{-1}g\circ R_{x_3,y_3}\circ g^{-1}gy\circ\iota_{1,2}\\
&= k(x,y)
\end{aligned}$$

Combining the k invariance with the $SO(V)$ invariance assumption of 12.5 gives the result:

$$\begin{aligned}(\mathcal{C}g \cdot s)(x) &= \frac{1}{|N_x|} \int\limits_{y \in N_x} k(x,y) \cdot s(g^{-1}y)\, \mathrm{d}\mu_o(y) \\ &= \frac{1}{|N_x|} \int\limits_{gy \in N_x} k_h(x,gy) \cdot s(y)\, \mathrm{d}\mu_o(y) \\ &= \frac{1}{|N_{g^{-1}x}|} \int\limits_{y \in N_{g^{-1}x}} k(g^{-1}x,y) \cdot s(y)\, \mathrm{d}\mu_o(y) \\ &= (g \cdot \mathcal{C}s)(x)\end{aligned}$$

□

Proof of theorem 13. For all $s \in W$ with $||s|| = 1$ we have:

$$s = \sum_{i=1}^{n} P_{V_i} s.$$

Applying the T quadratic form to s gives,

$$\begin{aligned}E \leq \langle Ts, s \rangle &= \sum_{i=1}^{n} \lambda_i \left|\left|P_{V_i} s\right|\right|^2 \\ &\leq \lambda_1 \sum_{i=1}^{m} \left|\left|P_{V_i} s\right|\right|^2 + \lambda_{m+1} \sum_{i=m+1}^{n} \left|\left|P_{V_i} s\right|\right|^2 \\ &= \lambda_1 (1 - \left|\left|P_{S_n^\perp} s\right|\right|^2) + \lambda_{m+1} \left|\left|P_{S_n^\perp} s\right|\right|^2.\end{aligned} \tag{C.9}$$

Thus,

$$\sqrt{\frac{\lambda_1 - E}{\lambda_1 - \lambda_{m+1}}} \geq \left|\left|P_{S_n}^\perp s\right|\right|$$

□

Appendix D

Charge transfer function

D.1 Definition of charge transfer function

Recall the idealized model for Cryo-EM images is given by a projection formula. Namely:

$$I(a,b) = \int_t \phi(ax(e_1) + bx(e_2) + tx(e_3))\,\mathrm{d}t,$$

for all $(a,b) \in \mathbb{R}^2$. Where I denotes the image, ϕ denotes the potential function, and x denotes the microscope orientation. Due to the physics of Cryo-EM image acquisition, recorded images from a Cryo-EM experiment are convolutions of these ideal projection images with a function φ which depends on the electron microscope, that is, observed images are given by:

$$I_{observed} = I_{projection} * \varphi.$$

In the Fourier domain we have:

$$\mathcal{F}(I_{observed}) = \mathcal{F}(I_{projection})\mathcal{F}(\varphi).$$

The function $\mathcal{F}(\varphi)$ is called the *charge transfer function* of the microscope and is abbreviated by CTF. It is approximated by a radial function given by:

$$\mathrm{CTF}(r) = \sin\left(-\pi\lambda\Delta z r^2 + \frac{\pi}{2}C_s\lambda^3 r^4 - \alpha\right)\exp(-\beta r^2),$$

for all $r \geq 0$. Where λ the wavelength of the electron, Δz is the defocus of the microscope, C_s is the spherical aberration, β is the Gaussian envelope decay,

and α is a constant. The defocus parameter Δz is of the most interest, because it can be changed easily in the experimental setup.

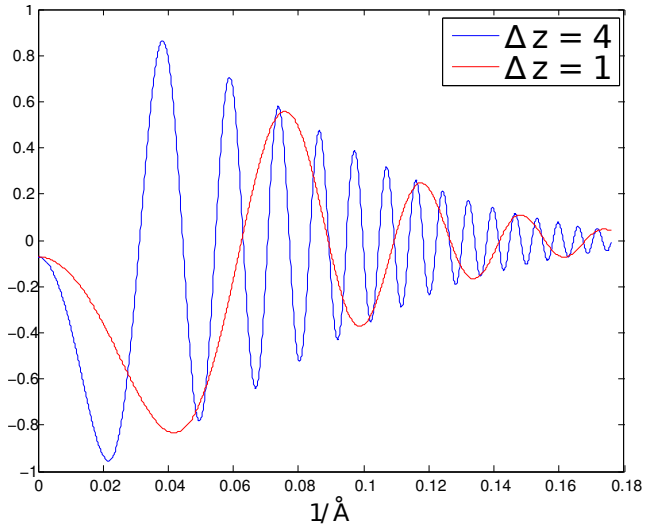


Figure D.1: CTFs with different defocus values. For the other constants: $C_s = 2\,\text{mm}$, $E = .0062\,\mu\text{m}$, $\beta = 100\,\mathring{A}$, and $\alpha = .07$.

D.2 Phase flipping and defocus groups

The CTF has two effects: it flips the sign of Fourier coefficients and dampens their magnitude. Notice the CTFs can have several zero crossings, therefore inverting the CTF is ill posed. A popular method to partially correct for the CTF is *phase flipping* which multiplies the Fourier coefficients of observed images by the sign of the CTF function:

$$\mathcal{F}(I_{flipped}) = \text{sgn}(\text{CTF})\mathcal{F}(I_{observed}) = |\text{CTF}|\mathcal{F}(I_{projection}).$$

The Fourier coefficients of the phase flipped images will have the correct phase, however the effect of Fourier dampening and zero crossings of the CTF are still present.

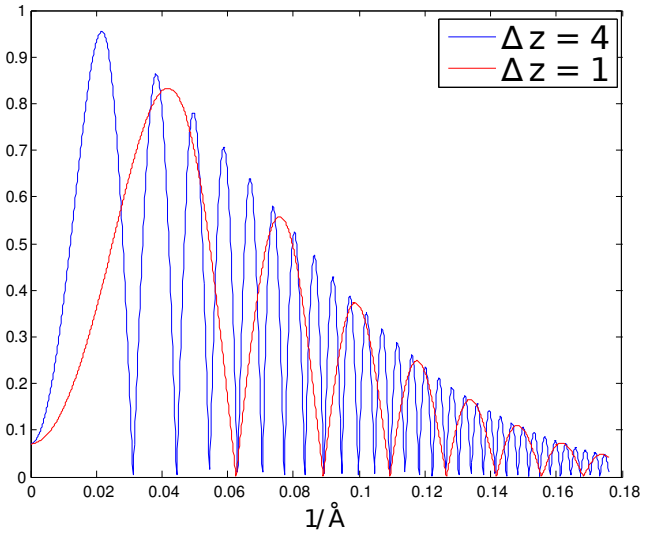


Figure D.2: Absolute value of CTFs from figure D.1. Phase flipped images are equal to $\mathcal{F}^{-1}\left(|\text{CTF}|\mathcal{F}(I_{projection})\right)$.

Images aquired using a single microscope defocus cannot be used to recover potential functions. This is because the image Fourier coefficients present at the zero crossing of the CTF are lost. To overcome this problem practitioners use *defocus groups*, that it, molecule samples are imaged using several different defocus values. The collection of images taken from a single defocus value is called a defocus group. Because the CTF for the different defocus settings have different zeros crossings, the defocus groups can be combined together to recover the Fourier coefficients of the potential function.

Bibliography

[1] Xiao-chen Bai, Israel S Fernandez, Greg McMullan, and Sjors HW Scheres. Ribosome structures to near-atomic resolution from thirty thousand cryo-em particles. *Elife*, 2:e00461, 2013.

[2] Alberto Bartesaghi, Doreen Matthies, Soojay Banerjee, Alan Merk, and Sriram Subramaniam. Structure of β-galactosidase at 3.2-å resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences*, 111(32):11709–11714, 2014.

[3] Ehsan Elhamifar and René Vidal. Sparse subspace clustering. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2790–2797. IEEE, 2009.

[4] Ehsan Elhamifar and René Vidal. Sparse manifold clustering and embedding. In J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 55–63. Curran Associates, Inc., 2011.

[5] Gerald B Folland. *A course in abstract harmonic analysis*. CRC press, 1994.

[6] Joachim Frank. Single-particle imaging of macromolecules by cryo-electron microscopy. *Annual review of biophysics and biomolecular structure*, 31:303–319, 2002.

[7] Jie Fu, Haixiao Gao, and Joachim Frank. Unsupervised classification of single particles by cluster tracking in multi-dimensional space. *Journal of Structural Biology*, 157(1):226–239, 2007.

[8] Robert M Glaeser. Electron crystallography of biological macromolecules. 2007.

[9] Robert M Glaeser and Richard J Hall. Reaching the information limit in cryo-em of biological macromolecules: experimental aspects. *Biophysical journal*, 100(10):2331–2337, 2011.

[10] Ronny Hadani and Amit Singer. Representation theoretic patterns in three dimensional Cryo-Electron Microscopy I - The intrinsic reconstitution algorithm. *Annals of Mathematics*, pages 1219–1241, 2009.

[11] Ronny Hadani and Amit Singer. Representation theoretic patterns in three-dimensional cryo-electron microscopy ii-the class averaging problem. *Foundations of Computational Mathematics*, 11(5):589–616, 2011.

[12] Ronny Hadani and Amit Singer. Representation theoretic patterns in three-dimensional cryo-electron microscopy ii—the class averaging problem. *Foundations of Computational Mathematics*, 11(5):589–616, 2011.

[13] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[14] Andres E Leschziner and Eva Nogales. Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions. *Annual review of biophysics and biomolecular structure*, 36:43–62, 2007.

[15] Xueming Li, Paul Mooney, Shawn Zheng, Christopher R Booth, Michael B Braunfeld, Sander Gubbens, David A Agard, and Yifan Cheng. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-em. *Nature methods*, 10(6):584–590, 2013.

[16] Maofu Liao, Erhu Cao, David Julius, and Yifan Cheng. Structure of the trpv1 ion channel determined by electron cryo-microscopy. *Nature*, 504(7478):107–112, 2013.

[17] Frank Natterer. *The mathematics of computerized tomography*, volume 32. Siam, 1986.

[18] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm.

[19] Pawel A Penczek. Chapter two-image restoration in cryo-electron microscopy. *Methods in enzymology*, 482:35–72, 2010.

[20] H Rullgård, L-G Öfverstedt, S Masich, B Daneholt, and Ozan Öktem. Simulation of transmission electron microscope images of biological specimens. *Journal of microscopy*, 243(3):234–256, 2011.

[21] Jeffrey Curtis Schlimmer. Concept acquisition through representational adjustment. 1987.

[22] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[23] Charles V Sindelar and Nikolaus Grigorieff. An adaptation of the wiener filter suitable for analyzing images of isolated single particles. *Journal of structural biology*, 176(1):60–74, 2011.

[24] Amit Singer, Zhizhen Zhao, Yoel Shkolnisky, and Ronny Hadani. Viewing angle classification of cryo-electron microscopy images using eigenvectors. *SIAM Journal on Imaging Sciences*, 4(2):723–759, 2011.

[25] S. M. Smith and J. M. Brady. Asset-2: real-time motion segmentation and shape tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):814–820, 1995.

[26] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

[27] Marin van Heel, Rodrigo Portugal, and Michael Schatz. Multivariate statistical analysis in single particle (cryo) electron microscopy. *Handbook on DVD 3D-EM in Life Sciences*, 2009.

[28] Marin van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of structural biology*, 151(3):250–262, 2005.

[29] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[30] Neil R Voss, Dmitry Lyumkis, Anchi Cheng, Pick-Wei Lau, Anke Mulder, Gabriel C Lander, Edward J Brignole, Denis Fellmann, Christopher Irving, Erica L Jacovetty, et al. A toolbox for ab initio 3-d reconstructions in single-particle electron microscopy. *Journal of structural biology*, 169(3):389–398, 2010.

[31] Lanhui Wang, Yoel Shkolnisky, and Amit Singer. A fourier-based approach for iterative 3d reconstruction from cryo-em images. *arXiv preprint arXiv:1307.5824*, 2013.

[32] Lanhui Wang, Amit Singer, and Zaiwen Wen. Orientation determination of cryo-em images using least unsquared deviations. *SIAM journal on imaging sciences*, 6(4):2450–2483, 2013.

[33] Qing Wang, Olaf Ronneberger, and Hans Burkhardt. Fourier analysis in polar and spherical coordinates. Technical report, Technical Report Internal Report 1/08, IIF-LMB, Computer Science Department, University of Freiburg, 2008.

[34] Zhao Wang, Corey F Hryc, Benjamin Bammes, Pavel V Afonine, Joanita Jakana, Dong-Hua Chen, Xiangan Liu, Matthew L Baker, Cheng Kao, Steven J Ludtke, et al. An atomic model of brome mosaic virus using direct electron detection and real-space optimization. *Nature communications*, 5, 2014.

[35] Norbert Wiener. Extrapolation, inerpolation, and smoothing of stationary time series. 1975.

[36] Zhizhen Zhao and Amit Singer. Fourier–bessel rotational invariant eigenimages. *JOSA A*, 30(5):871–877, 2013.

[37] Zhizhen Zhao and Amit Singer. Rotationally invariant image representation for viewing direction classification in cryo-emthen a. *Journal of structural biology*, 186(1):153–166, 2014.